

# CONTROLLED EQUILIBRIUM SELECTION IN STOCHASTICALLY PERTURBED DYNAMICS

ARI ARAPOSTATHIS, ANUP BISWAS, AND VIVEK S. BORKAR

**ABSTRACT.** We consider a dynamical system with finitely many equilibria and perturbed by small noise, in addition to being controlled by an ‘expensive’ control. The controlled process is optimal for an ergodic criterion with a running cost that consists the sum of the control effort and a penalty function on the state space. We study the optimal stationary distribution of the controlled process as the variance of the noise becomes vanishingly small. It is shown that depending on the relative magnitudes of the noise variance and the ‘running cost’ for control, one can identify three regimes, in each of which the optimal control forces the invariant distribution of the process to concentrate near equilibria that can be characterized according to the regime. We also obtain moment bounds for the optimal stationary distribution. Moreover, we show that in the vicinity of the points of concentration the density of optimal stationary distribution approximates the density of a Gaussian, and we explicitly solve for its covariance matrix.

## 1. INTRODUCTION

The study of dynamical systems has a long and profound history. A lot of effort has been devoted to understand the behavior of the system when it is perturbed by an additive noise [6, 17, 32]. Small noise diffusions have found applications in climate modeling [4, 5], electrical engineering [9, 43], finance [16] and many other areas. Recent work on ‘stochastic resonance’ (see, e.g., [31]) introduces an additional external input to the dynamics that may be viewed as a control. This is the main motivation for the study the model we introduce next.

**1.1. The model.** In this paper we consider a controlled dynamical system with small noise, which is modelled as a  $d$ -dimensional controlled diffusion  $X = [X_1, \dots, X_d]^T$  governed by the stochastic integral equation

$$X_t = X_0 + \int_0^t (m(X_s) + \varepsilon U_s) ds + \varepsilon^\nu W_t, \quad t \geq 0. \quad (1.1)$$

Here all processes live in a complete probability space  $(\Omega, \mathfrak{F}, \mathbb{P})$  and the data of (1.1) satisfies the following:

- (a)  $m = [m_1, \dots, m_d]^T : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a bounded  $\mathcal{C}^\infty$  function with bounded derivatives.
- (b)  $W$  is a standard Brownian motion in  $\mathbb{R}^d$ .

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING, THE UNIVERSITY OF TEXAS AT AUSTIN, 1 UNIVERSITY STATION, AUSTIN, TX 78712

DEPARTMENT OF MATHEMATICS, INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, DR. HOMI BHABHA ROAD, PUNE 411008, INDIA

DEPARTMENT OF ELECTRICAL ENGINEERING, INDIAN INSTITUTE OF TECHNOLOGY, POWAI, MUMBAI

*E-mail addresses:* ari@ece.utexas.edu, anup@iiserpune.ac.in, borkar@ee.iitb.ac.in.

*Date:* August 31, 2016.

2000 *Mathematics Subject Classification.* 35R60, 93E20.

*Key words and phrases.* Controlled diffusion, equilibrium selection, large deviations, small noise, expensive controls, ergodic control, HJB equation, ergodic LQG.

(c)  $U$  is an  $\mathbb{R}^d$ -valued control process which is jointly measurable in  $(t, \omega) \in [0, \infty) \times \bar{\Omega}$  (in particular it has measurable paths), and is *nonanticipative*: for  $t > s$ ,  $W_t - W_s$  is independent of

$$\mathfrak{F}_s := \text{the completion of } \sigma(X_0, W_r, U_r : r \leq s) \text{ relative to } (\mathfrak{F}, \mathbb{P}).$$

Such a control is called *admissible*, and we denote the set of admissible controls by  $\mathfrak{U}$ . As pointed out in [11, p. 18], we may, without loss of generality, assume that an admissible  $U$  is adapted to the natural filtration of  $X$ .

(d)  $0 < \varepsilon \ll 1$ .

(e)  $\nu > 0$ .

Let  $\mathcal{R}: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  be a *running cost* of the form

$$\mathcal{R}(x, u) := \ell(x) + \frac{1}{2} |u|^2, \quad (1.2)$$

where  $\ell: \mathbb{R}^d \rightarrow \mathbb{R}_+$  is a prescribed smooth, Lipschitz function satisfying the condition:

$$\lim_{|x| \rightarrow \infty} \ell(x) = \infty.$$

The control objective is to minimize the long run average (or *ergodic*) cost

$$\mathcal{J}(U) := \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \int_0^T \mathcal{R}(X_s, U_s) ds \right], \quad (1.3)$$

over all admissible controls. We define the *optimal value*  $\beta_*^\varepsilon$  by

$$\beta_*^\varepsilon := \inf_{U \in \mathfrak{U}} \mathcal{J}(U). \quad (1.4)$$

We view (1.1) as a perturbation of the o.d.e. (for *ordinary differential equation*)

$$\dot{x}(t) = m(x(t)), \quad (1.5)$$

perturbed by the ‘small noise’  $\varepsilon^\nu W_t$  (‘small’ because  $\varepsilon \ll 1$ ), and a control term  $\varepsilon U_t$ . Since  $\varepsilon$  is small, the optimization criterion in (1.3) implies that the control is ‘expensive’. We assume that the set of non-wandering points of the flow of (1.5) consists of finitely many hyperbolic equilibria, and that these are contained in some bounded open set which is positively invariant under the flow (see Hypothesis 1.1).

For the case when the control  $U \equiv 0$ , Freidlin and Wentzell developed a general framework for the analysis of small noise perturbed dynamical systems in [17] that is based on the theory of large deviations. Under a stochastic Lyapunov condition we introduce later (Hypothesis 1.1), the cost is finite for  $U = 0$ , ensuring in particular that the set of controls  $U \in \mathfrak{U}$  resulting in a finite value for  $\mathcal{J}(U)$  is nonempty. It is quite evident from ergodic theory that for  $U = 0$  the limit (1.3) is the expectation of  $\ell$  with respect to the invariant probability measure of (1.1).

The qualitative properties of the dynamics are best understood if we consider the special case  $d = 1$ , and  $m = -\frac{dF}{dx}$  for some smooth function  $F: \mathbb{R} \rightarrow \mathbb{R}$ . Then the trajectory of (1.5) converges to a critical point of  $F$ . In fact, generically (i.e., for  $x(0)$  in an open dense set) it converges to a stable one, i.e., to a local minimum. If one views the graph of  $F$  as a ‘landscape’, the local minima are the bottoms of its ‘valleys’. The behavior of the stochastically perturbed (albeit uncontrolled) version of this model, notably the analysis of where the stationary distribution concentrates, has been of considerable interest to physicists (see, e.g., [36, Chapter 8] or [17, Chapter 6]). To find the actual support of the limit in the case of multiple equilibria, one often looks at the large deviation properties of these invariant measures [17]. There are several studies in literature that deal with the large deviation principle of invariant measures of dynamical systems. Among the most relevant to the present are [14, 37] which obtain a large deviation principle for invariant measures (more precisely, invariant densities) of (1.1) under the assumption that there is a unique equilibrium point. This has been extended to multiple equilibria in [8]. A large deviation principle for invariant

measures for a class of reaction-diffusion systems is established in [13]. However, none of the above mentioned studies have any control component in their dynamics.

The model in (1.1) goes a step further and considers the full-fledged optimal control version of this, wherein one tries to induce a preferred equilibrium behavior through a feedback control. The reason the latter has to be ‘expensive’ is because this captures the physically realistic situation that one can ‘tweak’ the dynamics but cannot replace it by something altogether different without incurring considerable expense. The function  $\ell$  captures the relative preference among different points in the state space. Thus, the model in (1.1) is closely related to the model of stochastic resonance which has applications in neuron modelling, physics, electronics, physiology, etc. We refer to [19, Chapter 1] for various applications in the presence of small noise. In particular our model is closely related to the celebrated FitzHugh–Nagumo model [27] in the presence of noise. The control in (1.1) should be seen as an external input. In practice it is convenient to take  $U$  to be periodic in time, whereas we do not impose any periodicity constraint on  $U$ . The  $\varepsilon$  factor in the control could be interpreted as the *weak modulation* in [31]. We refer the reader to [31, 35] for a discussion on the interplay between noise variance and the control magnitude and its relation to stochastic resonance. Nonlinear control theory has been useful in understanding classes of systems that exhibit stochastic resonance [34]. Optimization theory has also been applied with the aim of enhancing the stochastic resonance effect for engineered systems [41, 42].

In our controlled setting we are interested in achieving a desired value of  $\beta_*^\varepsilon$ , reflecting the desired behavior of the corresponding stationary distribution. Although one can fix a suitable penalty function  $\ell$  beforehand, we will see in Theorem 1.11 in Section 1.3 that the value of  $\beta_*^\varepsilon$ , as well as the concentration of the stationary distribution change with  $\nu$ . Therefore a desired value of  $\beta_*^\varepsilon$  or a desired profile of the stationary distribution might be obtained for some specific values of  $\nu$  for small  $\varepsilon$ .

We also wish to point out that, since the control and noise are scaled differently, the ergodic control problem described can be viewed as a multi-scale diffusion problem.

**1.1.1. Assumptions on the vector field  $m$ .** Recall that a continuous-time dynamical system on a topological space  $\mathcal{X}$  is specified by a map  $\phi_t: \mathcal{X} \rightarrow \mathcal{X}$ , where  $\{\phi_t\}$  is a one parameter continuous abelian group action on  $\mathcal{X}$  called the *flow*. A point  $x \in \mathcal{X}$  is called *non-wandering* if for every open neighborhood  $U$  of  $x$  and every time  $T > 0$  there exists  $t > T$  such that  $\phi_t(U) \cap U \neq \emptyset$ .

Recall also that a critical point  $z$  of a smooth vector field  $m$  is called *hyperbolic* if the Jacobian matrix  $Dm(z)$  of  $m$  at  $z$  has no eigenvalues on the imaginary axis. For a hyperbolic critical point  $z$  of a vector field  $m$ , we let  $\mathcal{W}_s(z)$  and  $\mathcal{W}_u(z)$  denote the stable and unstable manifolds of its flow.

The following hypothesis on the vector field  $m$  is in effect throughout the paper.

**Hypothesis 1.1.** The vector field  $m$  is bounded and smooth and satisfies:

- (1) The set of non-wandering points of the flow of  $m$  is a finite set  $\mathcal{S} = \{z_1, \dots, z_n\}$  of hyperbolic critical points.
- (2) If  $y$  and  $z$  are critical points of  $m$ , then  $\mathcal{W}_s(y)$  and  $\mathcal{W}_u(z)$  intersect transversally (if they intersect).
- (3) There exist a smooth function  $\bar{\mathcal{V}}: \mathbb{R}^d \rightarrow \mathbb{R}_+$  and a bounded open neighborhood of the origin  $\mathcal{K} \subset \mathbb{R}^d$  containing  $S$ , with the following properties:
  - (3a)  $c_1|x|^2 \leq \bar{\mathcal{V}}(x) \leq c_2(1 + |x|^2)$  for some positive constants  $c_1, c_2$ , and all  $x \in \mathcal{K}^c$ .
  - (3b)  $\nabla \bar{\mathcal{V}}$  is Lipschitz and satisfies

$$\langle m(x), \nabla \bar{\mathcal{V}}(x) \rangle < -\gamma|x| \quad (1.6)$$

for some  $\gamma > 0$ , and all  $x \in \mathcal{K}^c$ .

**Remark 1.2.** The vector field  $m$  is assumed bounded for simplicity. The reader however might notice that the characterization of optimality (see Theorem 1.4) is based on the regularity results

in [3], and the hypotheses in [3, Section 4.6.1] permit  $m$  to be unbounded as long as

$$\limsup_{|x| \rightarrow \infty} \frac{|m(x)|^2}{\ell(x)} < \infty.$$

Provided that this condition is satisfied, the assumption that the drift is bounded can be waived and all the results of this paper hold unaltered, with the proofs requiring no major modification.

**1.1.2. Uncontrolled dynamics and quasi-potentials.** Recall the function  $\bar{\mathcal{V}}$  defined in Hypothesis 1.1. Since  $\nabla \bar{\mathcal{V}}$  is Lipschitz,  $\Delta \bar{\mathcal{V}}$  is bounded and thus (1.6) implies that with

$$\mathcal{L}_0^\varepsilon f(x) := \frac{\varepsilon^{2\nu}}{2} \Delta f(x) + \langle m(x), \nabla f(x) \rangle \quad \forall x \in \mathbb{R}^d, \quad f \in \mathcal{C}^2(\mathbb{R}^d),$$

we have

$$\mathcal{L}_0^\varepsilon \bar{\mathcal{V}}(x) \leq \gamma_0 - \gamma |x| \quad \forall \varepsilon \in (0, 1),$$

for some positive constants  $\gamma$  and  $\gamma_0$ . This Foster–Lyapunov condition implies in particular that the process  $X$  with  $U = 0$  has a unique invariant probability measure  $\eta_0^\varepsilon$ , and

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \int_0^T |X_t| dt \right] = \int_{\mathbb{R}^d} |x| \eta_0^\varepsilon(dx) \leq \frac{\gamma_0}{\gamma} \quad \forall \varepsilon \in (0, 1). \quad (1.7)$$

Since  $\ell$  is Lipschitz, (1.7) implies that there exists a constant  $\bar{c}_\ell$  independent of  $\varepsilon$  such that

$$\int \ell d\eta_0^\varepsilon \leq \bar{c}_\ell. \quad (1.8)$$

Moreover, from [8] there exists a unique Lipschitz continuous function  $Z \geq 0$ , such that  $\min_{\mathbb{R}^d} Z = 0$ ,  $Z(x) \rightarrow \infty$  as  $|x| \rightarrow \infty$  and

$$Z(x) = \inf_{\phi : \phi(t) \rightarrow x_i, x_i \in \mathcal{S}} \left[ \frac{1}{2} \int_0^\infty |\dot{\phi}(s) + m(\phi(s))|^2 ds + Z(x_i) \right], \quad \phi(0) = x,$$

and if  $\varrho_0^\varepsilon$  denotes the density of  $\eta_0^\varepsilon$ , then  $-\varepsilon^2 \ln \varrho_0^\varepsilon(x) \rightarrow Z(x)$  uniformly on compact subsets of  $\mathbb{R}^d$  as  $\varepsilon \searrow 0$ . The function  $Z$  is generally referred to as the *quasi-potential*, and plays a key role in the study of  $\eta_0^\varepsilon$ .

**1.2. The optimal stationary distribution.** For the model in (1.1) under the optimal control criterion in (1.3), the standard method of analysis using quasi-potentials no longer applies. The first important step is to characterize the stationary probability distributions of the controlled diffusion under optimal controls. It is evident that optimal controls belong to the class  $\widehat{\mathfrak{U}}$  defined by

$$\widehat{\mathfrak{U}} := \left\{ U \in \mathfrak{U} : \mathbb{E} \left[ \int_0^t |U_s|^2 ds \right] < \infty \text{ for all } t \geq 0 \right\}. \quad (1.9)$$

We state the following result concerning the existence of solutions to (1.1).

**Lemma 1.3.** *Under any  $U \in \widehat{\mathfrak{U}}$ , the diffusion in (1.1) has a unique strong solution.*

*Proof.* See Appendix A. □

In studying this problem, it is of course of paramount importance to assert the existence of an optimal stationary distribution, and ideally also prove that it is unique. A proper framework for this study is to consider the class of *infinitesimal ergodic occupation measures*, i.e., measures  $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$  which satisfy

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \mathcal{L}^\varepsilon[f](x, u) \pi(dx, du) = 0 \quad \forall f \in \mathcal{C}_c^\infty(\mathbb{R}^d), \quad (1.10)$$

where  $\mathcal{C}_c^\infty(\mathbb{R}^d)$  denotes the class of real-valued smooth functions with compact support. Here, the operator  $\mathcal{L}^\varepsilon: \mathcal{C}^2(\mathbb{R}^d) \rightarrow \mathcal{C}(\mathbb{R}^d \times \mathbb{R}^d)$  is defined by

$$\mathcal{L}^\varepsilon[f](x, u) := \frac{\varepsilon^{2\nu}}{2} \Delta f(x) + \langle m(x) + \varepsilon u, \nabla f(x) \rangle$$

for  $f \in \mathcal{C}^2(\mathbb{R}^d)$ . We adopt the usual relaxed control framework, where an admissible control is realized as a  $\mathcal{P}(\mathbb{R}^d)$ -valued measurable function (for details see [1, Section 2.3]). Thus if we disintegrate  $\pi$  as

$$\pi(dx, du) = \eta(dx) v(du | x),$$

and denote this as  $\pi = \eta \circledast v$ , then  $v$  is a relaxed Markov control, and  $\eta \in \mathcal{P}(\mathbb{R}^d)$  is an invariant probability measure for the corresponding controlled process, provided that the diffusion under the control  $v$  in (1.1) has a unique weak solution solution for all  $t \in [0, \infty)$  which is a Feller process.

Define

$$\mathcal{J}_\pi := \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathcal{R}(x, u) \pi(dx, du), \quad \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d).$$

Let  $\mathfrak{P}^\varepsilon$  denote the set of  $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$  satisfying (1.10), and consider the convex minimization problem

$$\mathcal{J}_* := \inf_{\pi \in \mathfrak{P}^\varepsilon} \mathcal{J}_\pi. \quad (1.11)$$

Recall that a function  $f: \mathbb{R}^m \rightarrow \mathbb{R}$  is called *inf-compact* if the set  $\{x \in \mathbb{R}^m : f(x) \leq C\}$  is compact (or empty) for every  $C \in \mathbb{R}$ . Since  $(x, u) \mapsto \ell(x) + \frac{1}{2}|u|^2$  is inf-compact (and the constraint is clearly feasible by (1.8)), and  $\pi \mapsto \mathcal{J}(\pi)$  is lower semi-continuous, it is straightforward to show that there exists some  $\pi_*^\varepsilon \in \mathfrak{P}^\varepsilon$  which attains the infimum (1.11). Also using the strict convexity of the map  $u \mapsto |u|^2$  one shows that  $\pi_*^\varepsilon$  is unique. However, it seems difficult to show that under the Markov control associated with  $\pi_*^\varepsilon$  the diffusion has a solution (See also Remark 1.7). For general results concerning this approach see [7, 26].

In the next section we adopt the dynamic programming approach to characterize the class of optimal stationary Markov controls and the corresponding invariant probability measures. Inevitably, we have to stay within a reasonably large class of controls under which the diffusion has a solution.

**1.2.1. Existence of an optimal control and the HJB equation.** Recall that a (precise) stationary Markov control is specified as  $U_t = v(X_t)$  for a measurable function  $v: \mathbb{R}^d \rightarrow \mathbb{R}^d$ . We identify the stationary Markov control with the function  $v$ . Let  $\mathfrak{U}_{SM}$  denote the class of stationary Markov controls which are locally bounded and under which (1.1) has a unique strong solution for all  $t \in [0, \infty)$ . Note that by [24, Theorem 2.5], under a control in  $\mathfrak{U}_{SM}$ , (1.1) has a unique solution up to explosion time. Moreover, it is strong Feller. Linear growth of  $|v|$  is the sufficient for the existence of a unique strong solution for all  $t \in [0, \infty)$ . We let  $\mathbb{E}_x^v$  denote the expectation operator on the canonical space of the process controlled by  $v \in \mathfrak{U}_{SM}$ , and starting at  $X_0 = x$ . We say that  $v \in \mathfrak{U}_{SM}$  is *stable* if the controlled process under  $v$  is positive recurrent, and we let  $\mathfrak{U}_{SSM}^\varepsilon \subset \mathfrak{U}_{SM}$  denote the set of *stable* controls in  $\mathfrak{U}_{SM}$ , and  $\mathcal{M}^\varepsilon$  the set of associated invariant probability measures. The following theorem essentially follows from [20, Theorem 2.2].

**Theorem 1.4.** *The HJB equation for the ergodic control problem, given by*

$$\frac{\varepsilon^{2\nu}}{2} \Delta V^\varepsilon + \min_{u \in \mathbb{R}^d} \left[ \langle m + \varepsilon u, \nabla V^\varepsilon \rangle + \ell + \frac{1}{2}|u|^2 \right] = \beta^\varepsilon, \quad (1.12)$$

*has no solution if  $\beta^\varepsilon > \beta_*^\varepsilon$ , while if  $\beta^\varepsilon < \beta_*^\varepsilon$  for any such solution  $V^\varepsilon$  the diffusion in (1.1) under the control  $v = -\varepsilon \nabla V^\varepsilon$  is transient. Moreover, the following hold:*

- (a) *If  $V^\varepsilon \in \mathcal{C}^2(\mathbb{R}^d)$  is any solution of (1.12), then  $|\nabla V^\varepsilon(x)|$  has at most affine growth in  $x$ ;*

- (b) If  $\beta^\varepsilon = \beta_*^\varepsilon$ , then (1.12) has a unique solution  $V^\varepsilon \in \mathcal{C}^2(\mathbb{R}^d)$  satisfying  $V^\varepsilon(0) = 0$ . The Markov control  $v_*^\varepsilon := -\varepsilon \nabla V^\varepsilon$  is stable, and is optimal for the ergodic control problem under the performance criterion in (1.3);
- (c) A control  $v \in \mathfrak{U}_{\text{SM}}$  is optimal only if it agrees with  $v_*^\varepsilon$  a.e. in  $\mathbb{R}^d$ .

*Proof.* The proof is contained in Appendix A.  $\square$

*Remark 1.5.* Due to the smoothness of coefficients, every weak solution in  $V^\varepsilon \in \mathcal{W}_{\text{loc}}^{1,\infty}(\mathbb{R}^d)$  of (1.12) is automatically in  $\mathcal{C}^k(\mathbb{R}^d)$  for any  $k \in \mathbb{N}$ . In the interest of notational economy, we often refer to any such  $V^\varepsilon$  as a solution, without specifying the function space it belongs to.

*Remark 1.6.* The existence and uniqueness result for the solution of (1.12) is well known [2,3] and in fact, the results in [3] hold for a more general class of HJB equations. However, we were not able to find any reference that establishes the verification of optimality results in Theorem 1.4. Recent work as in [21, 23] which investigates the optimal control problem, does not exactly fit our model. A strict growth condition for  $\ell$  is imposed in Assumption (H2) of [21], which we do not require here. On the other hand, in [23] where convergence of the Cauchy problem is investigated, and therefore optimality for the ergodic control problem is addressed, a more stringent condition is imposed (see Hypothesis (A3)') which for a Hamiltonian that is quadratic in the gradient like ours, amounts to geometric ergodicity under the uncontrolled dynamics.

The existence of a critical value for  $\beta^\varepsilon$  for (1.12) and the behavior of the solutions above or below this critical value are studied in detail in [20]. However, the critical value is not necessarily the optimal value in (1.4), and this is what the bulk of the proof in Appendix A is devoted to. For more recent work on the relation of the critical value of an elliptic HJB equation of the ergodic type and the optimal value of the control problem see [22].

For a stationary Markov control  $v$  we define the *extended generator* of (1.1) by

$$\mathcal{L}_v^\varepsilon f(x) := \frac{\varepsilon^{2\nu}}{2} \Delta f(x) + \langle m(x) + \varepsilon v(x), \nabla f(x) \rangle, \quad x \in \mathbb{R}^d, \quad (1.13)$$

for  $f \in \mathcal{C}^2(\mathbb{R}^d)$ .

From (1.12) it follows that

$$\frac{\varepsilon^{2\nu}}{2} \Delta V^\varepsilon + \langle m, \nabla V^\varepsilon \rangle - \frac{\varepsilon^2}{2} |\nabla V^\varepsilon|^2 + \ell = \beta_*^\varepsilon. \quad (1.14)$$

Theorem 1.4 shows that  $\beta_*^\varepsilon$  defined in (1.4) is attained at some  $v_*^\varepsilon \in \mathfrak{U}_{\text{SSM}}^\varepsilon$  and is independent of the initial condition  $X_0$ . Let  $\eta_*^\varepsilon$  denote the invariant probability measure of the diffusion under the control  $v_*^\varepsilon \in \mathfrak{U}_{\text{SSM}}$ . We have

$$\beta_*^\varepsilon = \int_{\mathbb{R}^d} \mathcal{R}(x, v_*^\varepsilon(x)) \eta_*^\varepsilon(dx) \quad (1.15)$$

by Birkhoff's ergodic theorem. Moreover, since any  $v \in \mathfrak{U}_{\text{SM}}$  which is optimal for the ergodic control problem agrees with  $v_*^\varepsilon$  a.e. in  $\mathbb{R}^d$ , it follows that  $\eta_*^\varepsilon$  is the unique probability measure in  $\mathcal{M}^\varepsilon$  which satisfies (1.15). Given this uniqueness property, we refer to  $\eta_*^\varepsilon$  as *the optimal invariant probability measure*, or as *the optimal stationary distribution*, and we let  $\varrho_*^\varepsilon$  denote its density. We also refer to  $v_*^\varepsilon$  as *the optimal stationary Markov control*.

*Remark 1.7.* We have restricted the class  $\mathfrak{U}_{\text{SM}}$  to locally bounded Markov controls for two reasons. First, it follows from the results in [18] that under any locally bounded Markov control which has at most affine growth, the diffusion in (1.1) has a unique strong solution. Second, for a locally bounded Markov control, any infinitesimal invariant probability measure, i.e., a probability measure  $\mu$  satisfying  $\int \mathcal{L}_v^\varepsilon f d\mu = 0$  for all  $f \in \mathcal{C}_c^\infty$ , has a positive continuous density. The positivity of the density is crucial in asserting the uniqueness (in an a.e. sense) of the optimal Markov control. However, more general statements of uniqueness are possible. As shown in [10] a sufficient condition

for the positivity of the density is that the drift in in the class  $L_{\text{loc}}^p(\mathbb{R}^d)$  for some  $p > d$ . On the other hand, by the results in [33], if  $|v| \in L_{\text{loc}}^p(\mathbb{R}^d)$  for  $p > d + 2$ , then (1.1) has a unique weak solution, which is moreover a Markov process. More recent results assert strong solutions and continuous dependence on the initial condition for locally integrable drifts [15, 24, 44]. Thus, the uniqueness result in Theorem 1.4(c) can be extended to a larger class of Markov controls which includes singular controls.

**1.3. Main results.** In this section we summarize the main results of the paper. We start with the following definition.

**Definition 1.8.** Let  $\mathcal{S}_s \subset \mathcal{S}$  denote the set of stable equilibria of (1.5), i.e., the set of points  $z \in \mathcal{S}$  for which the eigenvalues of  $Dm(z)$  have negative real parts.

We say that a set  $K \subset \mathbb{R}^d$  is *stochastically stable* (or that  $\eta_*^\varepsilon$  *concentrates* on  $K$ ) if it is compact, and  $\lim_{\varepsilon \searrow 0} \eta_*^\varepsilon(K^c) = 0$ . It is evident that the class  $\mathfrak{H}$  of stochastically stable sets, if nonempty, is closed under intersections. Hence we define the *minimal stochastically stable* set  $\mathfrak{S}$  by  $\mathfrak{S} := \cap_{K \in \mathfrak{H}} K$ .

The behavior of  $\eta_*^\varepsilon$  for small  $\varepsilon$  depends crucially on the parameter  $\nu$ . We distinguish three regimes: The *supercritical regime* ( $\nu > 1$ ), the *subcritical regime* ( $\nu < 1$ ), and the *critical regime* ( $\nu = 1$ ). Roughly speaking, the control ‘exceeds’ the noise level in the supercritical regime, while the opposite is the case in the subcritical regime. In the critical regime, which is the most interesting and more difficult to study, the control and noise levels are equal. The main results can be grouped in three categories: (1) characterization of the minimal stochastically stable set  $\mathfrak{S}$  and asymptotic estimates of  $\beta_*^\varepsilon$  for small  $\varepsilon$  in the three regimes (Theorem 1.11), (2) concentration bounds for  $\eta_*^\varepsilon$  (Theorem 1.12), and (3) convergence of  $\varrho_*^\varepsilon$ , under appropriate scaling, to a Gaussian density (Theorem 1.13).

**Definition 1.9.** For a square matrix  $M \in \mathbb{R}^{d \times d}$ , let  $\Lambda^+(M)$  denote the sum of its eigenvalues that lie in the open right half complex plane. For  $z \in \mathcal{S}$ , and with  $M_z := Dm(z)$ , where as defined earlier  $Dm(z)$  is the Jacobian of  $m$  at  $z$ , we let  $\widehat{Q}_z$  and  $\widehat{\Sigma}_z$  be the symmetric, nonnegative definite, square matrices solving the pair of equations

$$\begin{aligned} M_z^\top \widehat{Q}_z + \widehat{Q}_z M_z &= \widehat{Q}_z^2, \\ (M_z - \widehat{Q}_z) \widehat{\Sigma}_z + \widehat{\Sigma}_z (M_z - \widehat{Q}_z)^\top &= -I. \end{aligned} \tag{1.16}$$

By Theorem 1.18, which appears in Section 1.4, there exists a unique pair  $(\widehat{Q}_z, \widehat{\Sigma}_z)$  of symmetric positive semidefinite matrices solving (1.16). It is also evident by (1.16) that  $\widehat{\Sigma}_z$  is invertible.

Throughout the paper, the symbols  $\mathcal{O}(|x|^a)$  and  $\mathfrak{o}(|x|^a)$ , for  $a \in (0, \infty)$ , denote the sets of functions  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  having the property

$$\limsup_{|x| \searrow 0} \frac{|f(x)|}{|x|^a} < \infty, \quad \text{and} \quad \limsup_{|x| \searrow 0} \frac{|f(x)|}{|x|^a} = 0,$$

respectively. Abusing the notation,  $\mathcal{O}(|x|^a)$  and  $\mathfrak{o}(|x|^a)$  occasionally denote generic members of these sets. Thus, for example, an inequality of the form  $\mathcal{O}(|x|) \leq f(x) \leq \mathcal{O}(|x|)$  is well defined, and is equivalent to the statement that  $\limsup_{|x| \searrow 0} \frac{|f(x)|}{|x|} < \infty$ , and  $\liminf_{|x| \searrow 0} \frac{|f(x)|}{|x|^2} > -\infty$ .

In order to state the main results we need the following definition.

**Definition 1.10.** We define the *optimal control effort*  $\mathcal{G}_*^\varepsilon$  by

$$\mathcal{G}_*^\varepsilon := \frac{1}{2} \int_{\mathbb{R}^d} |v_*^\varepsilon|^2 d\eta_*^\varepsilon, \quad \varepsilon > 0. \tag{1.17}$$

Also define

$$\mathcal{Z}_c := \operatorname{Arg min}_{z \in \mathcal{S}} \{\ell(z) + \Lambda^+(Dm(z))\}, \quad \mathfrak{J}_c := \min_{z \in \mathcal{S}} [\ell(z) + \Lambda^+(Dm(z))],$$

$$\begin{aligned}
\mathcal{Z}_s &:= \operatorname{Arg\,min}_{z \in \mathcal{S}_s} \{\ell(z)\}, & \mathfrak{J}_s &:= \min_{z \in \mathcal{S}_s} [\ell(z)], \\
\mathcal{Z} &:= \operatorname{Arg\,min}_{z \in \mathcal{S}} \{\ell(z)\}, & \mathfrak{J} &:= \min_{z \in \mathcal{S}} [\ell(z)], \\
\tilde{\mathcal{Z}} &:= \operatorname{Arg\,min}_{z \in \mathcal{Z}} \{\Lambda^+(Dm(z))\}, & \tilde{\mathfrak{J}} &:= \min_{z \in \mathcal{Z}} [\Lambda^+(Dm(z))].
\end{aligned}$$

**Theorem 1.11.** *The minimal stochastically stable set  $\mathfrak{S}$  is a subset of  $\mathcal{S}$  for all  $\nu > 0$ . Also, the set  $\mathfrak{S}$ , the optimal value  $\beta_*^\varepsilon$ , and the optimal control effort  $\mathfrak{G}_*^\varepsilon$  depend on  $\nu$  as follows:*

(i) *For  $\nu > 1$  ('supercritical' regime),  $\mathfrak{S} \subset \tilde{\mathcal{Z}}$ , and*

$$\mathcal{O}(\varepsilon^{2\wedge\nu}) \leq \beta_*^\varepsilon - \mathfrak{J} \leq \mathcal{O}(\varepsilon^{2\nu}), \quad \mathfrak{G}_*^\varepsilon \in \mathcal{O}(\varepsilon^{\nu\wedge 2}) \quad \text{if } \mathfrak{J} = \mathfrak{J}_s,$$

$$\mathcal{O}(\varepsilon^{2\wedge\nu}) \leq \beta_*^\varepsilon - \mathfrak{J} \leq \varepsilon^{2\nu-2} \tilde{\mathfrak{J}} + \mathcal{O}(\varepsilon^{2\nu}), \quad \mathfrak{G}_*^\varepsilon \in \mathcal{O}(\varepsilon^{(2\nu-2)\wedge 2}) \quad \text{if } \mathfrak{J} < \mathfrak{J}_s.$$

(ii) *For  $\nu < 1$  ('subcritical' regime),  $\mathfrak{S} \subset \mathcal{Z}_s$ , and*

$$\mathcal{O}(\varepsilon^\nu) \leq \beta_*^\varepsilon - \mathfrak{J}_s \leq \mathcal{O}(\varepsilon^{\nu\vee(4\nu-2)}), \quad \mathfrak{G}_*^\varepsilon \in \mathcal{O}(\varepsilon^\nu). \quad (1.18)$$

(iii) *For  $\nu = 1$  ('critical' regime), we have  $\mathfrak{S} \subset \mathcal{Z}_c$ ,  $\beta_*^\varepsilon \leq \mathfrak{J}_c + \mathcal{O}(\varepsilon^2)$ , and  $\lim_{\varepsilon \searrow 0} \beta_*^\varepsilon = \mathfrak{J}_c$ . Moreover, if  $\mathfrak{J}_c = \mathfrak{J}_s$ , then the lower bound in (1.18) holds.*

It is not hard to show that the optimal invariant measures  $\eta_*^\varepsilon$  concentrate on  $\mathcal{S}$  as  $\varepsilon \searrow 0$  (see Lemma 3.1). In Theorem 1.11 we distinguish the three regimes corresponding to different values of  $\nu$ , and provide asymptotic bounds for  $\beta_*^\varepsilon$  for small  $\varepsilon$ . For  $\nu > 1$  one can find a control  $U$  under which the invariant measure of the dynamics (1.1) concentrates on a point in  $\mathcal{S}$ . Construction of invariant measures with similar properties is also possible for  $z \in \mathcal{S}_s$  when  $\nu < 1$ . The important difference is that for  $\nu < 1$  the optimal invariant measure  $\eta_*^\varepsilon$  cannot concentrate on  $\mathcal{S} \setminus \mathcal{S}_s$  (see Lemma 3.6). To show this fact we construct a suitable energy function for the Morse–Smale dynamics (see Theorem 2.2). The analysis in the critical regime  $\nu = 1$  turns out to be more subtle than the other two regimes. To facilitate the study of the critical regime, we identify an important property which concerns a singular ergodic control problem for Linear Quadratic Gaussian (LQG) systems (Theorem 1.18). This plays a crucial role in showing that  $\mathfrak{S} \subset \mathcal{Z}_c$ .

To guide the reader, we indicate the results presented in Sections 3–4 which comprise the proof of Theorem 1.11.

*Proof of Theorem 1.11.* That  $\mathfrak{S} \subset \mathcal{S}$  follows from Lemma 3.1. Part (i) follows from Lemma 3.3, Corollary 4.2, and Theorem 5.7. Part (ii) follows from Lemma 3.5 (ii), Lemma 3.6, and Corollary 4.6, and part (iii) follows from Lemma 3.3, Remark 4.7, and Theorem 5.4.  $\square$

The next theorem provides concentration bounds for the optimal stationary distribution in terms of moments. Let  $\operatorname{dist}(x, \mathcal{S})$  denote the Euclidean distance of  $x \in \mathbb{R}^d$  from the set  $\mathcal{S}$ , and  $B_r(\mathcal{S}) := \{y \in \mathbb{R}^d : \operatorname{dist}(y, \mathcal{S}) < r\}$ .

**Theorem 1.12.** *For any  $k \in \mathbb{N}$  and  $r > 0$ , there exist constants,  $\hat{\kappa}_0 = \hat{\kappa}_0(k, r, \nu)$ , and  $\hat{\kappa}_i = \hat{\kappa}_i(k)$ ,  $i = 1, 2$ , such that with  $\hat{r}(\varepsilon) := \hat{\kappa}_2 \varepsilon^{\nu\wedge 1}$  we have*

$$\int_{B_r(\mathcal{S})} (\operatorname{dist}(x, \mathcal{S}))^2 \eta_*^\varepsilon(dx) \leq \hat{\kappa}_0 \varepsilon^{2(\nu\wedge 2)} \quad \forall \nu > 0,$$

$$\int_{B_{\hat{r}(\varepsilon)}^c(\mathcal{S})} (\operatorname{dist}(x, \mathcal{S}))^{2k} \eta_*^\varepsilon(dx) \leq \hat{\kappa}_1 \varepsilon^{2(\nu\wedge 1)} \quad \forall \nu \in (0, 2],$$

for all  $\varepsilon \in (0, 1)$ .

Moreover, if  $D$  is any open set such that  $\mathcal{S}_s \subset D$ , then

$$\eta_*^\varepsilon(D^c) \in \mathcal{O}(\varepsilon^{2\nu\wedge(2-\nu)}),$$

provided  $\nu < 1$ , or  $\mathfrak{J}_c = \mathfrak{J}_s$  and  $\nu = 1$ , or  $\mathfrak{J} = \mathfrak{J}_s$  and  $\nu \in (1, 2)$ .

*Proof.* This follows by Lemma 4.1, Proposition 4.5, Corollary 4.6, and Remark 4.7.  $\square$

Exploiting the results in Theorem 1.12, we scale the space suitably and show that the resulting invariant measures are also tight. In particular, we examine the asymptotic behavior of  $\eta_*^\varepsilon$  and show that under an appropriate spatial scaling it ‘converges’ to a Gaussian distribution in the vicinity of the minimal stochastically stable set. This is the subject of the next theorem.

**Theorem 1.13.** *Assume  $\nu \in (0, 2)$ . Let  $z \in \mathcal{S}$ , and  $\mathcal{N}$  an open neighborhood of  $z$  whose closure does not contain any other elements of  $\mathcal{S}$ . Suppose that along some sequence  $\varepsilon_n \searrow 0$  we have  $\liminf_{\varepsilon_n \searrow 0} \eta_*^{\varepsilon_n}(\mathcal{N}) > 0$ . Then along this sequence it holds that*

$$\frac{\varepsilon^{\nu d} \varrho_*^\varepsilon(\varepsilon^\nu x + z)}{\eta_*^\varepsilon(\mathcal{N})} \xrightarrow{\varepsilon \searrow 0} \frac{1}{(2\pi)^{d/2} |\det \widehat{\Sigma}_z|^{1/2}} \exp\left(-\frac{1}{2} \langle x, \widehat{\Sigma}_z^{-1} x \rangle\right), \quad (1.19)$$

uniformly on compact sets, where ‘det’ denotes the determinant, and  $\widehat{\Sigma}_z$  is given by (1.16).

*Proof.* This follows from Theorems 5.3 and 5.7.  $\square$

We present a simple example to demonstrate the results.

**Example 1.14.** Let  $m$  be a vector field in  $\mathbb{R}$  of the form  $m = -\nabla F$ , with  $F$  a ‘double well potential’ given as follows:  $F(x) := \frac{x^4}{4} - \frac{x^3}{3} - x^2$  on  $[-10, 10]$ , with  $F$  suitably extended so that it is globally Lipschitz and does not have any critical points outside the interval  $[-10, 10]$ . Then  $\nabla F$  vanishes at exactly three points:  $-1, 0, 2$ . Of these, 0 is a local maximum, hence an unstable equilibrium for the o.d.e.  $\dot{x}(t) = m(x(t))$ , and both  $-1$  and  $2$  are local minima, hence stable equilibria thereof. Let  $\ell(x) = c|x|^2$  on  $[-10, 10]$  for a suitable  $c > 0$ , modified suitably outside  $[-10, 10]$  to render it globally Lipschitz. Note that  $F(0) = 0$ ,  $F(-1) = -\frac{5}{12}$ ,  $F(2) = -\frac{8}{3}$ . Thus  $x = 2$  is the unique global minimum of  $F$ . Since  $\ell(0) = 0$ , and  $Dm(0) = 2$ , the results of Theorem 1.11 indicate that:

- in the supercritical regime  $\mathfrak{S} = \{0\}$ , and  $\beta_*^\varepsilon \approx \ell(0) = 0$  for  $\varepsilon$  small;
- in the subcritical regime  $\mathfrak{S} = \{-1\}$ ,  $\beta_*^\varepsilon \approx \ell(-1) = c$  for  $\varepsilon$  small;
- in the critical regime, we have  $\mathfrak{S} = \{0\}$  if  $c > 2$ , with  $\beta_*^\varepsilon \approx \ell(0) + Dm(0) = 2$  for  $\varepsilon$  small, and  $\mathfrak{S} = \{-1\}$  if  $c < 2$ , with  $\beta_*^\varepsilon \approx \ell(-1) = c$  for  $\varepsilon$  small.

Next we change the data so that  $F(x) := \frac{x^6}{6} - \frac{x^5}{5} - \frac{7x^4}{4} + \frac{x^3}{3} + 3x^2$  on  $[-10, 10]$ . Then  $\nabla F$  vanishes at exactly five points, and  $\mathcal{S} = \{-2, -1, 0, 1, 3\}$ . Of these,  $-1$  and  $1$  are local maxima of  $F$ , hence unstable equilibria for the o.d.e.  $\dot{x}(t) = m(x(t))$ , while the rest are stable equilibria. Hence  $\mathcal{S}_s = \{-2, 0, 3\}$ . Let  $\ell(x) = 5x^4 - x^3 - 20x^2 + 16$  on  $[-10, 10]$ . The critical point  $z = 3$  is the unique global minimum for  $F$ , which means that it is stochastically stable for the uncontrolled dynamics. Calculating the values of  $\ell$  at  $\mathcal{S}$  we obtain:  $\ell(-2) = 24$ ,  $\ell(-1) = 2$ ,  $\ell(0) = 16$ ,  $\ell(1) = 0$ , and  $\ell(3) = 214$ . Also, we have  $Dm(-1) = 8$ ,  $Dm(1) = 12$ . By Theorem 1.11, we have:

- in the supercritical regime  $\mathfrak{S} = \{1\}$ , and  $\beta_*^\varepsilon \approx \ell(1) = 0$  for  $\varepsilon$  small;
- in the critical regime, we have  $\mathfrak{S} = \{-1\}$ , and  $\beta_*^\varepsilon \approx \ell(-1) + Dm(-1) = 10$  for  $\varepsilon$  small;
- in the subcritical regime  $\mathfrak{S} = \{0\}$ ,  $\beta_*^\varepsilon \approx \ell(0) = 16$  for  $\varepsilon$  small.

Note that in this example the stochastically stable sets are distinct in the three regimes.

**Remark 1.15.** Theorems 1.11–1.12 suggest that  $\nu = 2$  is a critical value. We present an example with linear drift and quadratic penalty, so that explicit calculations are possible, to show that indeed  $\nu = 2$  is a critical value. Consider a one-dimensional model with data  $m(x) = x$  and  $\ell(x) = (x+1)^2$ . Direct substitution shows that the solution of the HJB equation (see (1.14)) is

$$V^\varepsilon(x) = \frac{1 + \sqrt{1 + 2\varepsilon^2}}{2\varepsilon^2} \left( x + \frac{2\varepsilon^2}{(1 + \sqrt{1 + 2\varepsilon^2}) \sqrt{1 + 2\varepsilon^2}} \right)^2,$$

$$\beta_*^\varepsilon = \frac{1}{1+2\varepsilon^2} + \varepsilon^{2\nu-2} \frac{1+\sqrt{1+2\varepsilon^2}}{2}.$$

The closed loop drift is

$$\begin{aligned} x - \varepsilon^2 \nabla V^\varepsilon(x) &= -\sqrt{1+2\varepsilon^2}x - \frac{2\varepsilon^2}{\sqrt{1+2\varepsilon^2}} \\ &= -\sqrt{1+2\varepsilon^2} \left( x + \frac{2\varepsilon^2}{1+2\varepsilon^2} \right). \end{aligned} \quad (1.20)$$

Thus, the optimal stationary distribution  $\eta_*^\varepsilon$  is Gaussian with variance  $(\sigma_*^\varepsilon)^2$  and mean  $\mathfrak{m}_*^\varepsilon$  given by

$$(\sigma_*^\varepsilon)^2 := \frac{\varepsilon^{2\nu}}{2\sqrt{1+2\varepsilon^2}}, \quad \mathfrak{m}_*^\varepsilon := -\frac{2\varepsilon^2}{1+2\varepsilon^2}. \quad (1.21)$$

Consider the scaled distribution  $\hat{\eta}_*^\varepsilon$  with density  $\varepsilon^\nu \varrho_*^\varepsilon(\varepsilon^\nu x + z)$ . Let  $\mathcal{N}(\mathfrak{m}, \sigma^2)$  denote the Normal distribution with mean  $\mathfrak{m}$  and variance  $\sigma^2$ . We have

- For  $\nu \in (0, 2)$ ,  $\hat{\eta}_*^\varepsilon$  converges to  $\mathcal{N}(0, 1/2)$ .
- For  $\nu = 2$ ,  $\hat{\eta}_*^\varepsilon$  converges to  $\mathcal{N}(-2, 1/2)$ .
- For  $\nu > 2$ , we have  $\frac{\mathfrak{m}_*^\varepsilon}{\sigma_*^\varepsilon} \rightarrow -\infty$ , and thus  $\hat{\eta}_*^\varepsilon$  does not converge as  $\varepsilon \searrow 0$ .

Thus (1.19) does not hold for  $\nu \geq 2$ .

A simple calculation also shows that the optimal control effort is given by

$$\begin{aligned} \mathcal{G}_*^\varepsilon &= \frac{\varepsilon^{-2}}{2} \left( 1 + \sqrt{1+2\varepsilon^2} \right)^2 (\sigma_*^\varepsilon)^2 + \frac{\varepsilon^{-2}}{2} \left( 1 + \sqrt{1+2\varepsilon^2} \right)^2 \left( \frac{2\varepsilon^2}{(1+\sqrt{1+2\varepsilon^2})\sqrt{1+2\varepsilon^2}} + \mathfrak{m}_*^\varepsilon \right)^2 \\ &= \varepsilon^{2\nu-2} \frac{(1+\sqrt{1+2\varepsilon^2})^2}{4\sqrt{1+2\varepsilon^2}} + \frac{2\varepsilon^2}{(1+2\varepsilon^2)^2}. \end{aligned}$$

Thus  $\mathcal{G}_*^\varepsilon \in \mathcal{O}(\varepsilon^{(2\nu-2)\wedge 2})$  matching the analogous estimate in Theorem 1.11(i).

A better understanding of this can be reached by considering the limit  $\nu \rightarrow \infty$ , in which case the dynamics are deterministic. A simple calculation shows that

$$\bar{x} := \arg \min_x \left\{ \ell(\varepsilon x) + \frac{1}{2}|x|^2 \right\} = -\frac{2\varepsilon^2}{1+2\varepsilon^2}.$$

Thus for a feedback control to be optimal, the point  $\bar{x}$  should be asymptotically stable for the closed loop system. As a result, for the LQG problem, the optimal stationary distribution is centered at the point  $\bar{x}$  for all values of  $\nu$ . The criticality at  $\nu = 2$  is generic, since in the vicinity of an equilibrium  $z$ , solving the minimization problem we have  $\bar{x} \approx \varepsilon^2 \nabla \ell(z)$ .

There is a similar behavior if the drift is stable. Let  $m(x) = -x$ . We obtain

$$\begin{aligned} V^\varepsilon(x) &= \frac{-1+\sqrt{1+2\varepsilon^2}}{2\varepsilon^2} \left( x + \frac{2\varepsilon^2}{(-1+\sqrt{1+2\varepsilon^2})\sqrt{1+2\varepsilon^2}} \right)^2, \\ \beta_*^\varepsilon &= \frac{1}{1+2\varepsilon^2} + \varepsilon^{2\nu-2} \frac{-1+\sqrt{1+2\varepsilon^2}}{2} \\ &= 1 - \frac{2\varepsilon^2}{1+2\varepsilon^2} + \varepsilon^{2\nu} \frac{1}{1+\sqrt{1+2\varepsilon^2}}. \end{aligned}$$

The closed loop drift, variance, and mean are as in (1.20)–(1.21). Using the identity

$$\frac{-1+\sqrt{1+2\varepsilon^2}}{2\varepsilon^2} = \frac{1}{1+\sqrt{1+2\varepsilon^2}},$$

the optimal control effort takes the form

$$\begin{aligned}\mathcal{G}_*^\varepsilon &= \frac{2(\sigma_*^\varepsilon)^2}{(1 + \sqrt{1 + 2\varepsilon^2})^2} + \frac{2\varepsilon^2}{(1 + \sqrt{1 + 2\varepsilon^2})^2} \left( \frac{1 + \sqrt{1 + 2\varepsilon^2}}{\sqrt{1 + 2\varepsilon^2}} + \mathbf{m}_*^\varepsilon \right)^2 \\ &= \frac{\varepsilon^{2\nu}}{(1 + \sqrt{1 + 2\varepsilon^2})^2 \sqrt{1 + 2\varepsilon^2}} + \frac{2\varepsilon^2}{(1 + 2\varepsilon^2)^2}.\end{aligned}$$

So  $\mathcal{G}_*^\varepsilon \in \mathcal{O}(\varepsilon^{2\nu \wedge 2})$ .

The outline of the paper is as follows. In Section 1.4 we present an important property of LQG systems, which plays a crucial role in the study of the critical regime and also in the proof of Theorem 1.13. Section 1.5 summarizes the notation used in the paper. In Section 2 we discuss energy functions for gradient-like flows (Theorem 2.2). These are heavily used in the study of the subcritical regime. The proofs of the main results comprise Sections 3–5. Section 3 is devoted to the study of the minimal stochastically stable sets, Section 4 is primarily devoted to the proof of Theorem 1.12, while Section 5 studies the optimal stationary distribution under an appropriate scaling, which leads to Theorem 1.13. Appendix A contains the proofs of Lemma 1.3 and Theorem 1.4, while Appendix B is devoted to the proof of Lemma 1.16 and Theorem 1.18.

**1.4. A property of LQG systems.** As mentioned earlier, the study of the critical regime, and also the proof of Theorem 1.13 rely on an important property of LQG systems which we describe next. A matrix  $M \in \mathbb{R}^{d \times d}$  is called *exponentially dichotomous* if it has no eigenvalues on the imaginary axis. Consider the diffusion

$$dX_t = (MX_t + v(X_t)) dt + dW_t, \quad (1.22)$$

with  $M \in \mathbb{R}^{d \times d}$  exponentially dichotomous. Let  $\overline{\mathfrak{U}}_{\text{SSM}}$  denote the class of locally bounded stationary Markov controls  $v$ , under which the diffusion in (1.22) has a unique strong solution, is positive recurrent, and satisfies

$$\mathcal{E}(v) := \frac{1}{2} \int_{\mathbb{R}^d} |v(x)|^2 \mu_v(dx) < \infty, \quad (1.23)$$

where  $\mu_v$  denotes the associated invariant probability measure.

As Theorem 1.18 below asserts, the minimal control effort, defined by

$$\mathcal{E}_* := \inf_{v \in \overline{\mathfrak{U}}_{\text{SSM}}} \mathcal{E}(v),$$

which is required to render the diffusion positive recurrent by controls in  $\overline{\mathfrak{U}}_{\text{SSM}}$ , equals the trace of the unstable spectrum of the matrix  $M$ , which was denoted as  $\Lambda^+(M)$  in Definition 1.9. This result is related to classical results in deterministic linear control systems and the Riccati equation [25, 28, 40], but since we could not locate it in this form in the literature, a proof is included in Appendix B.

We start with the following lemma.

**Lemma 1.16.** *Provided  $M$  is exponentially dichotomous, there exists a constant  $\tilde{C}_0$  depending only on  $M$  such that*

$$\int_{\mathbb{R}^d} |x|^2 \mu_v(dx) \leq \tilde{C}_0 \left( 1 + \int_{\mathbb{R}^d} |v(x)|^2 \mu_v(dx) \right) \quad \forall v \in \overline{\mathfrak{U}}_{\text{SSM}}.$$

Recall that a real square matrix is called *Hurwitz* if its eigenvalues lie in the open left half complex plane. We need the following definition.

**Definition 1.17.** Let  $M \in \mathbb{R}^{d \times d}$  be fixed. Let  $\mathcal{G}(M)$  denote the collection of all matrices  $G \in \mathbb{R}^{d \times d}$  such that  $M - G$  is Hurwitz. For  $G \in \mathcal{G}(M)$ , let  $\Sigma_G$  denote the (unique) symmetric solution of the Lyapunov equation

$$(M - G)\Sigma_G + \Sigma_G(M - G)^T = -I, \quad (1.24)$$

and define

$$\begin{aligned} \mathcal{J}_G(M) &:= \frac{1}{2} \text{trace}(G\Sigma_G G^T), \\ \mathcal{J}_*(M) &:= \inf_{G \in \mathcal{G}(M)} \mathcal{J}_G(M). \end{aligned} \quad (1.25)$$

Let  $v_G(x) = -Gx$  for some  $G \in \mathbb{R}^{d \times d}$ . It is clear that for the diffusion in (1.22) to be positive recurrent under the linear control  $v_G$ , it is necessary that  $M - G$  be Hurwitz. If so, then the invariant probability distribution of the controlled diffusion is Gaussian with covariance matrix  $\Sigma_G$  given by (1.24). It is clear then that the control effort  $\mathcal{E}(v_G)$  defined in (1.23) satisfies  $\mathcal{E}(v_G) = \mathcal{J}_G(M)$ . Therefore, provided the infimum in (1.25) is attained, then  $\mathcal{J}_*(M)$  is the minimal control effort, as defined by (1.23), required to render (1.22) positive recurrent using a linear stationary Markov control. Theorem 1.18 asserts that the infimum in (1.25) is indeed attained and that  $\mathcal{J}_*(M) = \Lambda^+(M)$ . Moreover, linear stationary Markov controls are optimal for this task within the class  $\bar{\mathfrak{U}}_{\text{SSM}}$ .

**Theorem 1.18.** Suppose that  $M \in \mathbb{R}^{d \times d}$  is exponentially dichotomous. Then the following hold:

- (a) There exists a unique positive semidefinite symmetric solution  $Q$  of the matrix Riccati equation

$$M^T Q + Q M = Q^2, \quad (1.26)$$

satisfying

$$(M - Q)\Sigma + \Sigma(M - Q)^T = -I \quad (1.27)$$

for some symmetric positive definite matrix  $\Sigma$ . Moreover,  $A = M - Q$  attains the infimum in (1.25) subject to (1.24), and it holds that

$$\mathcal{J}_*(M) = \Lambda^+(M) = \frac{1}{2} \text{trace}(Q).$$

- (b) With  $\mu_v$  denoting the invariant probability measure of (1.22) under a control  $v \in \bar{\mathfrak{U}}_{\text{SSM}}$ , we have

$$\inf_{v \in \bar{\mathfrak{U}}_{\text{SSM}}} \int_{\mathbb{R}^d} \frac{1}{2}|v(x)|^2 \mu_v(dx) = \Lambda^+(M). \quad (1.28)$$

Moreover, any control  $v_* \in \bar{\mathfrak{U}}_{\text{SSM}}$  which attains the infimum in (1.28) satisfies  $v_*(x) = -Qx$  for almost all  $x$  in  $\mathbb{R}^d$ .

- (c) Let  $\bar{\beta} \in \mathbb{R}$ . The equation

$$\frac{1}{2} \Delta \bar{V}(x) + \langle Mx, \nabla \bar{V}(x) \rangle - \frac{|\nabla \bar{V}(x)|^2}{2} = \bar{\beta} \quad (1.29)$$

has no solution if  $\bar{\beta} > \Lambda^+(M)$ . If  $\bar{\beta} = \Lambda^+(M)$ , then  $\bar{V}(x) = \frac{1}{2}\langle x, Qx \rangle$  is the unique solution of (1.29) satisfying  $\bar{V}(0) = 0$ . If  $\bar{\beta} < \Lambda^+(M)$  and  $\bar{V}$  is a solution of (1.29), then the diffusion in (1.22) under the control  $v = -\nabla \bar{V}$  is transient.

**Remark 1.19.** Optimality and uniqueness of the optimal control  $v(x) = -Qx$  in Theorem 1.18 (b) holds over a larger class of Markov controls. Indeed combining the results of [10, 24], we can replace ‘locally bounded’ in the definition of  $\bar{\mathfrak{U}}_{\text{SSM}}$  by  $|\hat{v}^\varepsilon| \in L_{\text{loc}}^p(\mathbb{R}^d)$  for some  $p > d$ . Then the results of Theorem 1.18 (b) hold for this class of controls.

**1.5. Notation.** The following notation is used in this paper. The symbols  $\mathbb{R}$ , and  $\mathbb{C}$  denote the fields of real numbers, and complex numbers, respectively. Also,  $\mathbb{N}$  denotes the set of natural numbers. The Euclidean norm on  $\mathbb{R}^d$  is denoted by  $|\cdot|$ , and  $\langle \cdot, \cdot \rangle$  denotes the inner product. For two real numbers  $a$  and  $b$ ,  $a \wedge b := \min(a, b)$  and  $a \vee b := \max(a, b)$ . For a matrix  $M$ ,  $M^\top$  denotes its transpose, and  $\|M\|$  denotes the operator norm relative to the Euclidean vector norm. Also  $I$  denotes the identity matrix.

The composition of two functions  $f$  and  $g$  is denoted by  $f \circ g$ . A ball of radius  $r > 0$  in  $\mathbb{R}^d$  around a point  $x$  is denoted by  $B_r(x)$ , or as  $B_r$  if  $x = 0$ . For a compact set  $K$ , we let  $\text{dist}(x, K)$  denote the Euclidean distance of  $x \in \mathbb{R}^d$  from the set  $K$ , and  $B_r(K) := \{y \in \mathbb{R}^d : \text{dist}(y, K) < r\}$ . For a set  $A \subset \mathbb{R}^d$ , we use  $\bar{A}$ ,  $A^c$ , and  $\partial A$  to denote the closure, the complement, and the boundary of  $A$ , respectively. We define  $\mathcal{C}_b^k(\mathbb{R}^d)$ ,  $k \geq 0$ , as the set of functions whose  $i^{\text{th}}$  derivatives,  $i = 1, \dots, k$ , are continuous and bounded in  $\mathbb{R}^d$  and denote by  $\mathcal{C}_c^k(\mathbb{R}^d)$  the subset of  $\mathcal{C}_b^k(\mathbb{R}^d)$  with compact support. The space of all probability measures on a Polish space  $\mathcal{X}$  with the Prohorov topology is denoted by  $\mathcal{P}(\mathcal{X})$ . The density of the  $d$ -dimensional Gaussian distribution with mean 0 and covariance matrix  $\Sigma$  is denoted by  $\rho_\Sigma$ .

The term *domain* in  $\mathbb{R}^d$  refers to a nonempty, connected open subset of the Euclidean space  $\mathbb{R}^d$ . We introduce the following notation for spaces of real-valued functions on a domain  $G \subset \mathbb{R}^d$ . The space  $L^p(G)$ ,  $p \in [1, \infty)$ , stands for the usual Banach space of (equivalence classes of) measurable functions  $f$  satisfying  $\int_G |f(x)|^p dx < \infty$ , and  $L^\infty(G)$  is the Banach space of functions that are essentially bounded in  $G$ . The standard Sobolev space of functions on  $G$  whose generalized derivatives up to order  $k$  are in  $L^p(G)$ , equipped with its natural norm, is denoted by  $\mathcal{W}^{k,p}(G)$ ,  $k \geq 0$ ,  $p \geq 1$ .

In general if  $\mathcal{Y}$  is a space of real-valued functions on a domain  $G$ ,  $\mathcal{Y}_{\text{loc}}$  consists of all functions  $f$  such that  $f\varphi \in \mathcal{Y}$  for every  $\varphi \in \mathcal{C}_c^\infty(G)$ , the space of smooth functions on  $G$  with compact support. In this manner we obtain for example the space  $\mathcal{W}_{\text{loc}}^{2,p}(G)$ .

Also  $\kappa_1, \kappa_2, \dots$  are generic constants whose definition differs from place to place.

## 2. GRADIENT-LIKE FLOWS AND ENERGY FUNCTIONS

**2.1. Gradient-Like Morse–Smale dynamical systems.** It is well known from the theory of dynamical systems that if the set of non-wandering points of a flow on a compact manifold consists of hyperbolic fixed points, then the associated vector field is generically *gradient-like* (see Definition 2.1 and Theorem 2.2 below). This is also the case under Hypothesis 1.1, since the ‘point at infinity’ is a source for the flow of  $m$ .

Recall that the *index* of a hyperbolic critical point  $z \in \mathbb{R}^d$  of a smooth vector field is defined as the dimension of the unstable manifold  $\mathcal{W}_u(z)$ . This agrees with the number of eigenvalues of  $Dm(z)$  which have positive real parts. The theorem below is well known [30, 38]. What we have added in its statement is the assertion that the energy function can be chosen in a manner that its Laplacian at critical points of the vector field with positive index is negative.

We start with the following definition.

**Definition 2.1.** We say that  $\mathcal{V} \in \mathcal{C}^\infty(\mathbb{R}^d)$  is an *energy function* if it is inf-compact, and has a finite set  $\mathcal{S} = \{z_1, \dots, z_n\}$  of critical points, which are all nondegenerate. A  $\mathcal{C}^\infty$  vector field  $m$  on  $\mathbb{R}^d$  is called *gradient-like relative to* an energy function  $\mathcal{V}$  provided that the set of non-wandering points of its flow is  $\mathcal{S}$ , that every point in  $\mathcal{S}$  is a hyperbolic critical point of  $m$ , and

$$\langle m(x), \nabla \mathcal{V}(x) \rangle < 0 \quad \forall x \in \mathbb{R}^d \setminus \mathcal{S}.$$

If  $m$  satisfies these properties, we also say that  $m$  is *adapted to*  $\mathcal{V}$ .

**Theorem 2.2.** Suppose that  $m$  is a smooth vector field in  $\mathbb{R}^d$  for which Hypothesis 1.1 holds. Let  $G$  be any domain of  $\mathbb{R}^d$  of the form  $\{x \in \mathbb{R}^d : \bar{\mathcal{V}} < c\}$  for some  $c \in \mathbb{R}$ , satisfying  $G \supset \mathcal{K}$ , and let  $\{a_z : z \in \mathcal{S}\}$  be any set of distinct real numbers such that if  $z$  and  $z'$  are the  $\alpha$ - and  $\omega$ -limit points

of some trajectory, respectively, then  $a_z > a_{z'}$ . Then there exists a function  $\widehat{\mathcal{V}} \in \mathcal{C}^\infty(\bar{G})$ , with the following properties:

- (i)  $\langle m(x), \nabla \widehat{\mathcal{V}}(x) \rangle < 0$  for all  $x \in \bar{G} \setminus \mathcal{S}$ .
- (ii) For each  $z \in \mathcal{S}$ , there exists a neighborhood  $\mathcal{N}_z$  of  $z$  and a symmetric matrix  $Q_z \in \mathbb{R}^{d \times d}$  such that  $\widehat{\mathcal{V}}(x) = a_z + \langle x - z, Q_z(x - z) \rangle + \mathcal{O}(|x - z|^2)$  for all  $x \in \mathcal{N}_z$ .
- (iii)  $\Delta \widehat{\mathcal{V}}(z) < 0$ , for all  $z \in \mathcal{S} \setminus \mathcal{S}_s$ , where  $\mathcal{S}_s$ , as defined earlier, denotes the stable equilibria of the flow of  $m$ .
- (iv) There exists a constant  $C_0 > 0$  such that

$$C_0 (\text{dist}(x, \mathcal{S}) \vee |\nabla \widehat{\mathcal{V}}(x)|)^2 \leq |\langle m(x), \nabla \widehat{\mathcal{V}}(x) \rangle| \leq C_0^{-1} (\text{dist}(x, \mathcal{S}) \wedge |\nabla \widehat{\mathcal{V}}(x)|)^2 \quad (2.1)$$

for all  $x \in G$ .

*Proof.* Since  $m$  is smooth and bounded, and  $m(z) = 0$  for  $z \in \mathcal{S}$ , there exists a constant  $\tilde{C}_m > 0$  such that

$$|M_z x - m(x)| \leq \tilde{C}_m |x|^2 \quad \forall x \in \mathbb{R}^d, \quad \forall z \in \mathcal{S}. \quad (2.2)$$

Let  $z \in \mathcal{S}$  be a critical point of  $m$  of index  $q \geq 0$ . Translating the coordinates we may assume that  $z = 0$ . Since  $m(0) = 0$ , then by (2.2),  $m(x)$  takes the form

$$m(x) = Mx + \mathcal{O}(|x|^2)$$

locally around  $x = 0$ , where  $M = Dm(0)$ . By hypothesis  $M$  has exactly  $q$  ( $d - q$ ) eigenvalues in the open right half (left half) complex space. Therefore since the corresponding eigenspaces are invariant under  $M$ , there exists a linear coordinate transformation  $T$  such that, in the new coordinates  $\tilde{x} = T(x)$ , the linear map  $x \mapsto Mx$  has the matrix representation  $\tilde{M} = TMT^{-1}$  and  $\tilde{M} = \text{diag}(\tilde{M}_1, -\tilde{M}_2)$ , where  $\tilde{M}_1$  and  $\tilde{M}_2$  are square Hurwitz matrices of dimension  $d - q$  and  $q$  respectively. By the Lyapunov theorem there exist positive definite matrices  $\tilde{Q}_i$ ,  $i = 1, 2$ , satisfying

$$\begin{aligned} \tilde{M}_1^\top \tilde{Q}_1 + \tilde{Q}_1 \tilde{M}_1 &= -I_{d-q}, \\ \tilde{M}_2^\top \tilde{Q}_2 + \tilde{Q}_2 \tilde{M}_2 &= -I_q, \end{aligned} \quad (2.3)$$

where  $I_{d-q}$  and  $I_q$  are the identity matrices of dimension  $d - q$  and  $q$ , respectively. Suppose  $q > 0$ , and let  $\theta > 1$  be such that

$$\theta \text{trace}(T^\top \text{diag}(0, \tilde{Q}_2)T) > \text{trace}(T^\top \text{diag}(\tilde{Q}_1, 0)T), \quad (2.4)$$

and define  $\widehat{\mathcal{V}}$  in some neighborhood of 0 by

$$\widehat{\mathcal{V}}(x) := a + \langle x, T^\top \text{diag}(\tilde{Q}_1, -\theta \tilde{Q}_2)Tx \rangle, \quad (2.5)$$

where  $a$  is a constant to be determined later. By (2.4) we obtain  $\Delta \widehat{\mathcal{V}}(0) < 0$ , and thus (ii) holds.

Using (2.2), we have

$$\langle m(x), \nabla \widehat{\mathcal{V}}(x) \rangle = x^\top [M^\top T^\top \text{diag}(\tilde{Q}_1, -\theta \tilde{Q}_2)T + T^\top \text{diag}(\tilde{Q}_1, -\theta \tilde{Q}_2)TM]x + \mathcal{O}(|x|^3).$$

Expanding we obtain

$$\begin{aligned} T^\top \text{diag}(\tilde{Q}_1, -\theta \tilde{Q}_2)TM &= T^\top \text{diag}(\tilde{Q}_1, -\theta \tilde{Q}_2)TT^{-1}\tilde{M}T \\ &= T^\top \text{diag}(\tilde{Q}_1 \tilde{M}_1, \theta \tilde{Q}_2 \tilde{M}_2)T. \end{aligned}$$

By (2.3) we obtain

$$\langle m(x), \nabla \widehat{\mathcal{V}}(x) \rangle = -\langle x, T^\top \text{diag}(I_{d-q}, \theta I_q)Tx \rangle + \mathcal{O}(|x|^3).$$

Therefore, since  $\theta > 1$ , we have

$$-\|Tx\|^2 + \mathcal{O}(|x|^3) \leq \langle m(x), \nabla \widehat{\mathcal{V}}(x) \rangle \leq -\theta \|Tx\|^2 + \mathcal{O}(|x|^3). \quad (2.6)$$

As shown in [38] one can select any real numbers  $a_i$  and define  $\widehat{\mathcal{V}}$  on  $\mathcal{S}$  by setting  $\widehat{\mathcal{V}}(z_i) = a_i$  as long as the following consistency condition is met: if  $z_i$  and  $z_j$  are the  $\alpha$ - and  $\omega$ -limit points of some trajectory then  $a_i > a_j$ . Thus  $\widehat{\mathcal{V}}$  can be defined in non-overlapping neighborhoods of the critical points by (2.5) so as to satisfy (2.6) and parts (i)–(iii) of the theorem. Since  $G$  is positively invariant under the flow of  $m$ , the stable and unstable manifolds of  $\mathcal{S}$  intersect transversally by Hypothesis 1.1 (2), and  $m$  is transversal to the boundary of  $\partial G$  by Hypothesis 1.1 (3b), this function can then be extended to  $\bar{G}$  by the handlebody decomposition technique introduced by Smale. For details see [38, Theorem B] and [30, Theorem 1].

It is clear by (2.5)–(2.6) that (2.1) holds in some open neighborhood of each  $z \in \mathcal{S}$ , and thus,  $\mathcal{S}$  being a finite set, it also holds in some neighborhood of  $\mathcal{N}$  of  $\mathcal{S}$ . Since  $\langle m, \nabla \widehat{\mathcal{V}} \rangle$  is strictly negative on the compact set  $\bar{G} \setminus \mathcal{N}$  and  $\langle m(x), \nabla \widehat{\mathcal{V}}(x) \rangle < 0$  for all  $x \notin \mathcal{S}$ , a constant  $C_0$  can be selected so that (2.1) holds on  $G$ . This completes the proof.  $\square$

The function  $\widehat{\mathcal{V}}$  in Theorem 2.2 can be extended to  $\mathbb{R}^d$ , and constructed in a manner so that it agrees, outside some ball, with the Lyapunov function  $\bar{\mathcal{V}}$  in Hypothesis 1.1. This is stated in the following lemma.

**Lemma 2.3.** *Under the assumptions of Theorem 2.2 the vector field  $m$  is adapted to an energy function  $\mathcal{V}$  which satisfies  $\mathcal{V} = \bar{\mathcal{V}}$  on the complement of some open ball which contains  $\mathcal{S}$ . Also parts (i)–(iv) of Theorem 2.2 hold, and for every bounded domain  $G$  there exists a constant  $C_0 = C_0(G)$  such that (2.1) holds for all  $x \in G$ . Moreover there exists a constant  $\bar{C}_0 > 0$  such that with*

$$\begin{aligned}\bar{\mathcal{V}}(x) &:= \max \left\{ (\text{dist}(x, \mathcal{S}))^2 \wedge \text{dist}(x, \mathcal{S}), |\nabla \mathcal{V}(x)|^2 \wedge |\nabla \mathcal{V}(x)| \right\}, \\ \underline{\mathcal{V}}(x) &:= \min \left\{ (\text{dist}(x, \mathcal{S}))^2 \wedge \text{dist}(x, \mathcal{S}), |\nabla \mathcal{V}(x)|^2 \wedge |\nabla \mathcal{V}(x)| \right\},\end{aligned}$$

we have

$$(\bar{C}_0)^{-1} \bar{\mathcal{V}}(x) \leq |\langle m(x), \nabla \mathcal{V}(x) \rangle| \leq \bar{C}_0 \underline{\mathcal{V}}(x) \quad \forall x \in \mathbb{R}^d. \quad (2.7)$$

*Proof.* Select  $c \in \mathbb{R}$  such that  $G_1 := \{x \in \mathbb{R}^d : \bar{\mathcal{V}} < c\}$  contains  $\mathcal{K}$ . Let  $G_2 := \{x \in \mathbb{R}^d : \bar{\mathcal{V}} < 2c\}$ . By Theorem 2.2 there exists  $\widehat{\mathcal{V}} \in \mathcal{C}^\infty(G_2)$  with the properties stated. Without loss of generality we can assume that  $\widehat{\mathcal{V}} = 2c$  on  $\partial G_2$  [38, Theorem B]. Let  $c_1 := \sup_{G_1} \widehat{\mathcal{V}}$ . Then  $c_1 < 2c$  by the positive invariance of  $G_2$ , and the property  $\langle m, \nabla \widehat{\mathcal{V}} \rangle < 0$  in  $G_2 \setminus G_1$ . We write  $A \Subset B$  to indicate that  $\bar{A} \subset B$ . Let  $\tilde{G} := \{x \in \mathbb{R}^d : \widehat{\mathcal{V}} < (c_1 + 2c)/2\}$ , and  $c_2 := \sup_{\tilde{G}} \bar{\mathcal{V}}$ . Then  $G_1 \Subset \tilde{G} \Subset G_2$ , and  $c < c_2 < 2c$  by construction.

Let  $\psi: \mathbb{R} \rightarrow \mathbb{R}$  be a smooth non-decreasing function such that  $\psi(t) = t$  for  $t \leq \frac{1}{2}(c_1 + 2c)$ ,  $\psi(t) = 2c$  for  $t \geq 2c$ , and whose derivative is strictly positive on the interval  $[\frac{1}{2}(c_1 + 2c), 2c]$ . Similarly, let  $\bar{\psi}: \mathbb{R} \rightarrow \mathbb{R}$  be a smooth non-decreasing function such that  $\bar{\psi}(t) = 0$  for  $t \leq -c$  and  $\bar{\psi}(t) = t$  for  $t \geq c_2 - 2c$ . Define  $\mathcal{V} := \psi \circ \widehat{\mathcal{V}} + \bar{\psi} \circ (\bar{\mathcal{V}} - 2c)$ . By construction  $\mathcal{V}$  agrees with  $\widehat{\mathcal{V}}$  on  $G_1$  and with  $\bar{\mathcal{V}}$  on  $G_2^c$ . It can also be easily verified that  $\sup_{G_2 \setminus G_1} \langle m, \nabla \mathcal{V} \rangle < 0$ . Thus  $\mathcal{V} \in \mathcal{C}^\infty(\mathbb{R}^d)$  is an energy function, and  $m$  is adapted to  $\mathcal{V}$  according to Definition 2.1.

Since  $\langle m(x), \nabla \mathcal{V}(x) \rangle < 0$  for all  $x \notin \mathcal{S}$ , and  $\mathcal{V}$  agrees with  $\widehat{\mathcal{V}}$  on  $\mathcal{K}$ , Theorem 2.2 (i)–(iv) clearly hold. Also since (2.7) holds in some neighborhood of  $\mathcal{S}$  by (2.5)–(2.6), then, in view of the linear growth of  $\langle m(x), \nabla \bar{\mathcal{V}}(x) \rangle \neq 0$  in (1.6), and the assumptions on the growth of  $\bar{\mathcal{V}}$  in Hypothesis 1.1, (2.7) also holds on  $\mathbb{R}^d$ .  $\square$

### 3. MINIMAL STOCHASTICALLY STABLE SETS

Recall that  $\beta_*^\varepsilon$  denotes the optimal value of (1.3),  $\eta_*^\varepsilon$  denotes the stationary distribution of the process  $X$  under the optimal stationary Markov control  $v_*^\varepsilon$ , and  $\varrho_*^\varepsilon$  denotes its density. These definitions are fixed throughout the rest of the paper. Also recall the definition of the extended

generator in (1.13), and the definition of  $\mathcal{R}$  in (1.2). For a stationary Markov control  $v$ , we use the notation

$$\mathcal{R}[v](x) := \mathcal{R}(x, v(x)) = \ell(x) + \frac{1}{2} |v(x)|^2. \quad (3.1)$$

Throughout the rest of the paper  $\mathcal{V}$  is a smooth function that satisfies (i)–(iv) in Theorem 2.2 and agrees with  $\bar{\mathcal{V}}$  in Hypothesis 1.1 on the complement of some open ball which contains  $\mathcal{S}$  (Lemma 2.3). We refer to  $\mathcal{V}$  as the *energy function*.

We start the analysis with the following lemma which asserts that  $\eta_*^\varepsilon$  concentrates on  $\mathcal{S}$  as  $\varepsilon \searrow 0$ .

**Lemma 3.1.** *The family  $\{\eta_*^\varepsilon, \varepsilon \in (0, 1)\}$  is tight, and any sub-sequential limit as  $\varepsilon \searrow 0$  has support on  $\mathcal{S}$ .*

*Proof.* Recall that  $\eta_0^\varepsilon$  denotes the invariant probability measure of (1.1) under the control  $U = 0$ . Define

$$\beta_0^\varepsilon := \int_{\mathbb{R}^d} \ell(x) \eta_0^\varepsilon(dx).$$

By (1.8) we have

$$\int_{\mathbb{R}^d} \ell(x) \eta_*^\varepsilon(dx) \leq \beta_*^\varepsilon \leq \beta_0^\varepsilon \leq \bar{c}_\ell \quad \forall \varepsilon \in (0, 1). \quad (3.2)$$

Since  $\ell$  is inf-compact, (3.2) implies that  $\{\eta_*^\varepsilon, \varepsilon \in (0, 1)\}$  is tight. Let  $\phi_t(x)$  denote the solution of (1.5) starting at  $x \in \mathbb{R}^d$  at  $t = 0$ , i.e.,  $\phi_0(x) = x$ . If  $C_m$  denotes a Lipschitz constant of  $m$  and  $X_0 = x$ , we have

$$|X_t - \phi_t(x)| \leq C_m \int_0^t |X_s - \phi_s(x)| ds + \varepsilon \int_0^t |v_*^\varepsilon(X_s)| ds + \varepsilon^\nu |W_t|. \quad (3.3)$$

Hence applying Gronwall's inequality we obtain from (3.3) that

$$\sup_{s \in [0, t]} |X_s - \phi_s(x)| \leq e^{C_m t} \left( \varepsilon \int_0^t |v_*^\varepsilon(X_s)| ds + \varepsilon^\nu \sup_{s \leq t} |W_s| \right). \quad (3.4)$$

In turn, for any  $\delta > 0$ , (3.4) implies that

$$\mathbb{P}_x(|X_t - \phi_t(x)| \geq \delta) \leq \mathbb{P}_x \left( \int_0^t |v_*^\varepsilon(X_s)| ds \geq \frac{\delta e^{-C_m t}}{2\varepsilon} \right) + \mathbb{P}_x \left( \sup_{s \leq t} |W_s| \geq \frac{\delta e^{-C_m t}}{2\varepsilon^\nu} \right)$$

for  $t > 0$ . By Jensen's inequality we obtain

$$\begin{aligned} \mathbb{P}_x \left( \int_0^t |v_*^\varepsilon(X_s)| ds \geq \frac{\delta e^{-C_m t}}{2\varepsilon} \right) &\leq \mathbb{P}_x \left( \int_0^t |v_*^\varepsilon(X_s)|^2 ds \geq \frac{\delta^2 e^{-2C_m t}}{4t\varepsilon^2} \right) \\ &\leq \frac{4t\varepsilon^2}{\delta^2} e^{2C_m t} \mathbb{E}_x \left[ \int_0^t |v_*^\varepsilon(X_s)|^2 ds \right]. \end{aligned}$$

Therefore for any compact set  $K \subset \mathbb{R}^d$  we have

$$\begin{aligned} \int_K \mathbb{P}_x(|X_t - \phi_t(x)| \geq \delta) \eta_*^\varepsilon(dx) &\leq \frac{4t^2 \varepsilon^2}{\delta^2} e^{2C_m t} \int_{\mathbb{R}^d} |v_*^\varepsilon(x)|^2 \eta_*^\varepsilon(dx) \\ &\quad + \sup_{x \in K} \mathbb{P}_x \left( \sup_{s \leq t} |W_s| \geq \frac{\delta}{2\varepsilon^\nu} e^{-C_m t} \right). \end{aligned} \quad (3.5)$$

It is clear that the right hand side of (3.5) tends to 0 as  $\varepsilon \searrow 0$ . Thus for any compact set  $K \subset \mathbb{R}^d$ , and any Lipschitz function  $f \in \mathcal{C}_b(\mathbb{R}^d)$  it holds that

$$\int_K |\mathbb{E}_x^{v_*^\varepsilon}[f(X_t)] - f(\phi_t(x))| \eta_*^\varepsilon(dx) \xrightarrow[\varepsilon \searrow 0]{} 0. \quad (3.6)$$

On the other hand, since  $\eta_*^\varepsilon$  is an invariant probability measure, we have

$$\int_{\mathbb{R}^d} \mathbb{E}_x^{v_*^\varepsilon}[f(X_t)] \eta_*^\varepsilon(dx) = \int_{\mathbb{R}^d} f(x) \eta_*^\varepsilon(dx) \quad \forall f \in \mathcal{C}_b(\mathbb{R}^d), \quad \forall t \geq 0. \quad (3.7)$$

Let  $\bar{\eta} \in \mathcal{P}(\mathbb{R}^d)$  be any limit of  $\eta_*^\varepsilon$  along some sequence  $\{\varepsilon_n\}$ , with  $\varepsilon_n \searrow 0$  as  $n \rightarrow \infty$ . By (3.6)–(3.7), the tightness of  $\{\eta_*^\varepsilon, \varepsilon \in (0, 1)\}$ , and a standard triangle inequality, we obtain

$$\int_{\mathbb{R}^d} f(\phi_t(x)) \bar{\eta}(dx) = \int_{\mathbb{R}^d} f(x) \bar{\eta}(dx) \quad \forall t \geq 0, \quad (3.8)$$

for all Lipschitz functions  $f \in \mathcal{C}_b(\mathbb{R}^d)$ . Since the  $\omega$ -limit set of any trajectory of (1.5) is contained in  $\mathcal{S}$ , (3.8) shows that  $\bar{\eta}$  has support on  $\mathcal{S}$ . This completes the proof.  $\square$

**3.1. Two Lemmas concerning the case  $\nu \geq 1$ .** For  $z \in \mathcal{S}$ , let  $\bar{v}_z^\varepsilon$ ,  $\varepsilon \in (0, 1)$ , denote the stationary Markov control defined by

$$\bar{v}_z^\varepsilon(x) := \frac{(M_z - \hat{Q}_z)(x - z) - m(x)}{\varepsilon}, \quad t \geq 0, \quad (3.9)$$

where  $M_z$  and  $\hat{Q}_z$  are as in Definition 1.9. The controlled process, is then governed by the diffusion

$$dX_t = (M_z - \hat{Q}_z)(X_t - z) dt + \varepsilon^\nu dW_t. \quad (3.10)$$

Since  $M_z - \hat{Q}_z$  is Hurwitz by Theorem 1.18, the diffusion has a stationary probability distribution  $\bar{\mu}_z^\varepsilon$ , which is Gaussian with mean  $z$  and covariance matrix  $\varepsilon^{2\nu} \hat{\Sigma}_z$ , where  $\hat{\Sigma}_z$  is as in (1.16).

We start with the following lemma.

**Lemma 3.2.** *Suppose that  $\nu \geq 1$  and  $z \in \mathcal{S}$ . Let  $\bar{v}_z^\varepsilon$  be the stationary Markov control in (3.9), and  $\bar{\mu}_z^\varepsilon$  the invariant probability measure of the diffusion governed by (3.10). Then*

$$\begin{aligned} \int_{\mathbb{R}^d} \frac{1}{2} |\bar{v}_z^\varepsilon(x)|^2 \bar{\mu}_z^\varepsilon(dx) &= \varepsilon^{2\nu-2} \Lambda^+(Dm(z)) + \mathcal{O}(\varepsilon^{4\nu-2}), \\ \int_{\mathbb{R}^d} \ell(x) \bar{\mu}_z^\varepsilon(dx) &= \ell(z) + \mathcal{O}(\varepsilon^{2\nu}). \end{aligned} \quad (3.11)$$

*Proof.* Without loss of generality assume that  $z = 0$ , and simplifying the notation we let  $M = M_z$ ,  $Q = \hat{Q}_z$ ,  $\Sigma = \hat{\Sigma}_z$ , and  $\bar{\mu}^\varepsilon = \bar{\mu}_z^\varepsilon$ .

We have

$$|(M - Q)x - m(x)|^2 = |Qx|^2 + 2\langle Qx, Mx - m(x) \rangle + |Mx - m(x)|^2. \quad (3.12)$$

Since by Taylor's theorem it holds that

$$\langle Qx, Mx - m(x) \rangle = \langle Qx, F(x) \rangle + \mathcal{O}(|x|^4),$$

with

$$F(x) := (F_1(x), \dots, F_d(x)) \quad \text{and} \quad F_i(x) := \frac{1}{2} \langle x, \nabla^2 m_i(0)x \rangle,$$

by (2.2) and (3.12) we obtain

$$|(M - Q)x - m(x)|^2 = |Qx|^2 + 2\langle Qx, F(x) \rangle + \mathcal{O}(|x|^4). \quad (3.13)$$

As mentioned in the paragraph preceding the lemma,  $\bar{\mu}^\varepsilon$  is Gaussian, with zero mean, and covariance matrix  $\varepsilon^{2\nu} \Sigma$ , where  $\Sigma$  is the solution of (1.27). Since  $\langle Qx, F(x) \rangle$  is a homogeneous polynomial of degree 3 it has zero mean under the Gaussian. Also the fourth moments of  $\bar{\mu}^\varepsilon$  are of order  $\varepsilon^{4\nu}$ . It then follows by the estimate in (3.13) and Theorem 1.18 (b) that

$$\begin{aligned} \frac{1}{2} \int_{\mathbb{R}^d} |\bar{v}_z^\varepsilon(x)|^2 \bar{\mu}_z^\varepsilon(dx) &= \int_{\mathbb{R}^d} \frac{1}{2\varepsilon^2} |Qx|^2 \bar{\mu}_z^\varepsilon(dx) + \mathcal{O}(\varepsilon^{4\nu-2}) \\ &= \varepsilon^{2\nu-2} \Lambda^+(M) + \mathcal{O}(\varepsilon^{4\nu-2}). \end{aligned} \quad (3.14)$$

To prove the second equation in (3.11), we use the bound

$$|\ell(x) - \ell(z) - D\ell(z)(x - z)| \leq \tilde{C}_\ell |x - z|^2 \quad \forall x \in \mathbb{R}^d, \quad \forall z \in \mathcal{S}, \quad (3.15)$$

for some constant  $\tilde{C}_\ell$ , and since  $\bar{\mu}^\varepsilon$  has zero mean we obtain

$$\left| \int_{\mathbb{R}^d} (\ell(x) - \ell(0)) \bar{\mu}^\varepsilon(dx) \right| \leq \varepsilon^{2\nu} \tilde{C}_\ell \operatorname{trace}(\Sigma). \quad (3.16)$$

By combining (3.14) and (3.16) we obtain (3.11). This completes the proof.  $\square$

Recall the notation in Definition 1.10. Lemma 3.2 in conjunction with Lemma 3.1 leads to:

**Lemma 3.3.** *It holds that*

$$\begin{aligned} \beta_*^\varepsilon &\leq \mathfrak{J} + \varepsilon^{2\nu-2} \min_{z \in \mathcal{Z}} \Lambda^+(Dm(z)) + \mathcal{O}(\varepsilon^{2\nu}) \quad \text{if } \nu > 1, \\ \beta_*^\varepsilon &\leq \mathfrak{J}_c + \mathcal{O}(\varepsilon^2) \quad \text{if } \nu = 1. \end{aligned} \quad (3.17)$$

Moreover, if  $\nu > 1$ , then

$$\lim_{\varepsilon \searrow 0} \beta_*^\varepsilon = \mathfrak{J}, \quad (3.18)$$

and  $\mathfrak{S} \subset \mathcal{Z}$ .

*Proof.* Recall the function  $\mathcal{R}$  defined in (3.1). Since

$$\beta_*^\varepsilon \leq \int_{\mathbb{R}^d} \mathcal{R}[\bar{v}^\varepsilon](x) \bar{\mu}_z^\varepsilon(dx) \quad \forall z \in \mathcal{S},$$

the first inequality in (3.17) follows by evaluating (3.11) at a point  $z \in \tilde{\mathcal{Z}}$ , while the second inequality in (3.17) follows by evaluating (3.11) at a point  $z \in \mathcal{Z}_c$ .

Since

$$\lim_{\varepsilon \searrow 0} \beta_*^\varepsilon \geq \mathfrak{J} \quad (3.19)$$

for all  $\nu > 0$  by Lemma 3.1, (3.18) follows by (3.17) and (3.19) when  $\nu > 1$ , and clearly then, in this case we have  $\mathfrak{S} \subset \mathcal{Z}$ .  $\square$

*Remark 3.4.* It is worth mentioning here that if  $z \in \mathcal{S}_s$ , then a control that renders  $\{z\}$  stochastically stable can be synthesized from the energy function  $\mathcal{V}$ . Note that by Theorem 2.2 (ii),  $\mathcal{V}$  can be selected so that  $\mathcal{V}(z) = 0$  and  $\mathcal{V}(z') > 0$  for all  $z' \in \mathcal{S} \setminus \{z\}$ . Consider the control

$$\check{v}^\varepsilon(x) := -\frac{1}{\varepsilon}(m(x) + \nabla \mathcal{V}(x)), \quad t \geq 0.$$

Then  $X$  is given by

$$dX_t = -\nabla \mathcal{V}(X_t) dt + \varepsilon^\nu dW_t, \quad t \geq 0.$$

Let  $\check{\mu}^\varepsilon$  denote its unique invariant probability measure. Recall the definition in (1.13). Since

$$\mathcal{L}_{\check{v}^\varepsilon}^\varepsilon \mathcal{V} \leq \frac{\varepsilon^{2\nu}}{2} \|\Delta \mathcal{V}\|_\infty - |\nabla \mathcal{V}|^2,$$

it follows that

$$2 \int_{\mathbb{R}^d} |\nabla \mathcal{V}|^2 d\check{\mu}^\varepsilon \leq \varepsilon^{2\nu} \|\Delta \mathcal{V}\|_\infty.$$

Note that  $\check{\mu}^\varepsilon$  has density  $\varrho^\varepsilon(x) = C(\varepsilon) e^{-\frac{2\mathcal{V}(x)}{\varepsilon^{2\nu}}}$ , where  $C(\varepsilon)$  is a normalizing constant. Therefore we have

$$\begin{aligned} \int_{\mathbb{R}^d} |\check{v}^\varepsilon(x)|^2 \check{\mu}^\varepsilon(dx) &\leq 2 \int_{\mathbb{R}^d} (|m(x)|^2 + |\nabla \mathcal{V}(x)|^2) \varepsilon^{-2} \check{\mu}^\varepsilon(dx) \\ &\leq 2 \int_{\mathbb{R}^d} \varepsilon^{-2} |m(x)|^2 \check{\mu}^\varepsilon(dx) + \varepsilon^{2\nu-2} \|\Delta \mathcal{V}\|_\infty \end{aligned}$$

$$\leq \mathcal{O}(\varepsilon^{2\nu-2}) + \varepsilon^{2\nu-2} \|\Delta\mathcal{V}\|_\infty.$$

For the last inequality we used the fact that  $m$  is bounded,  $m(z) = 0$ , and that  $\mathcal{V}$  is locally quadratic around  $z$ .

**3.2. Results concerning stable equilibria.** Recall that  $\mathcal{S}_s$  is the collection of stable equilibrium points, and  $\mathfrak{J}_s = \min_{z \in \mathcal{S}_s} \{\ell(z)\}$ . The following lemma holds for any  $\nu > 0$ . It shows that if  $z \in \mathcal{S}_s$  then there exists a Markov stationary control  $v^\varepsilon$  with invariant measure  $\mu^\varepsilon$  satisfying  $\int_{\mathbb{R}^d} |v^\varepsilon(x)|^2 \mu^\varepsilon(dx) \in \mathcal{O}(\varepsilon^n)$  for any  $n \in \mathbb{N}$ , under which  $\{z\}$  is stochastically stable.

**Lemma 3.5.** *The following hold.*

(i) *For any  $\nu > 0$  and  $z \in \mathcal{S}_s$  there exists a Markov control  $\check{v}^\varepsilon$ , and constants  $\varepsilon_0 = \varepsilon_0(\nu) > 0$ , and  $c_0 > 0$  independent of  $\nu$ , with the following properties. With  $\check{\mu}^\varepsilon$  denoting the invariant probability measure of (1.1) under the control  $\check{v}^\varepsilon$ , it holds that*

$$\begin{aligned} \int_{|x-z| \geq \varepsilon^{\nu/2}} |x-z|^2 \check{\mu}^\varepsilon(dx) &\leq \frac{\varepsilon^{2\nu}}{c_0(1-\varepsilon^\nu)} e^{-c_0\varepsilon^{-\nu}}, \\ \int_{\mathbb{R}^d} |\check{v}^\varepsilon(x)|^2 \check{\mu}^\varepsilon(dx) &\leq \frac{\varepsilon^{2(\nu-1)}}{c_0(1-\varepsilon^\nu)} e^{-c_0\varepsilon^{-\nu}} \end{aligned} \quad (3.20)$$

for all  $\varepsilon < \varepsilon_0$ , and

$$\varepsilon^{-\nu} \left| \int_{\mathbb{R}^d} \ell(x) \check{\mu}^\varepsilon(dx) - \ell(z) \right| \xrightarrow[\varepsilon \searrow 0]{} 0. \quad (3.21)$$

In particular, we have

$$\limsup_{\varepsilon \searrow 0} \frac{1}{\varepsilon^n} \int_{\mathbb{R}^d} |\check{v}^\varepsilon(x)|^2 \check{\mu}^\varepsilon(dx) = 0 \quad \forall n \in \mathbb{N}.$$

(ii) *It holds that  $\beta_*^\varepsilon \leq \mathfrak{J}_s + o(\varepsilon^\nu)$  for  $\nu \in (0, 2/3)$ , and  $\beta_*^\varepsilon \leq \mathfrak{J}_s + \mathcal{O}(\varepsilon^{4\nu-2})$  for  $\nu \in [2/3, 1)$ .*

*Proof.* In order to simplify the notation, we translate the origin so that  $z = 0$ , and we let  $M := Dm(0)$ . Let  $R^{-1}$  be the symmetric positive definite solution to the Lyapunov equation  $MR^{-1} + R^{-1}M^\top = -4I$ . Thus  $M^\top R + RM = -4R^2$ . Since scaling  $R$  by multiplying it with a positive constant smaller than 1 preserves the inequality

$$M^\top R + RM \leq -4R^2, \quad (3.22)$$

we may assume that  $\text{trace}(R) \leq 1$  and (3.22) holds. The sole purpose of this scaling is to simplify the calculations in the proof. We define the control  $\check{v}^\varepsilon$  by

$$\check{v}^\varepsilon(x) := \begin{cases} \varepsilon^{-1}(Mx - m(x)) & \text{if } |Rx| \geq \varepsilon^{\nu/2}, \\ 0 & \text{otherwise.} \end{cases}$$

We apply the function  $F(x) := \varepsilon^{2\nu} \exp(\varepsilon^{-2\nu} \langle x, Rx \rangle)$  to  $\mathcal{L}_{\check{v}^\varepsilon}^\varepsilon$ , which is defined in (1.13). By (3.22), and since  $\text{trace}(R) \leq 1$ , we obtain

$$\begin{aligned} \mathcal{L}_{\check{v}^\varepsilon}^\varepsilon F(x) &= (\varepsilon^{2\nu} \text{trace}(R) + 2|Rx|^2 + \langle x, (M^\top R + RM)x \rangle) e^{\frac{\langle x, Rx \rangle}{\varepsilon^{2\nu}}} \\ &\leq (\varepsilon^{2\nu} - 2|Rx|^2) e^{\frac{\langle x, Rx \rangle}{\varepsilon^{2\nu}}} \quad \text{if } |Rx| \geq \varepsilon^{\nu/2}. \end{aligned} \quad (3.23)$$

If  $|Rx| < \varepsilon^{\nu/2}$ , then  $\check{v}^\varepsilon = 0$ , and we obtain

$$\begin{aligned} \mathcal{L}_{\check{v}^\varepsilon}^\varepsilon F(x) &= (\varepsilon^{2\nu} \text{trace}(R) + 2|Rx|^2 + 2\langle m(x), Rx \rangle) e^{\frac{\langle x, Rx \rangle}{\varepsilon^{2\nu}}} \\ &\leq (\varepsilon^{2\nu} - 2|Rx|^2 + 2|Mx - m(x)||Rx|) e^{\frac{\langle x, Rx \rangle}{\varepsilon^{2\nu}}} \end{aligned}$$

$$\leq (\varepsilon^{2\nu} - |Rx|^2) e^{\frac{\langle x, Rx \rangle}{\varepsilon^{2\nu}}} \quad \text{for } |Rx| < \varepsilon^{\nu/2} \wedge \frac{1}{2}\|R\|^2 \tilde{C}_m^{-1}, \quad (3.24)$$

where in the first inequality we use (3.22), and in the second we use (2.2). Thus selecting  $\varepsilon_0$  as

$$\varepsilon_0 := 1 \wedge \left( \frac{1}{2}\|R\|^2 \tilde{C}_m^{-1} \right)^{2/\nu},$$

then, provided  $\varepsilon < \varepsilon_0$ , (3.24) holds for all  $x$  such that  $|Rx| < \varepsilon^{\nu/2}$ . It follows by (3.23) and (3.24) that  $\mathcal{L}_{\check{v}^\varepsilon}^\varepsilon F(x) \leq 0$  if  $|Rx| \geq \varepsilon^\nu$ , and

$$\sup \left\{ \mathcal{L}_{\check{v}^\varepsilon}^\varepsilon F(x) : |Rx| \leq \varepsilon^\nu, \varepsilon < \varepsilon_0 \right\} \leq e^{\|R^{-1}\| \varepsilon^{2\nu}} \quad \forall \varepsilon < \varepsilon_0. \quad (3.25)$$

Thus, by (3.23), (3.24), and (3.25), we obtain

$$\mathcal{L}_{\check{v}^\varepsilon}^\varepsilon F(x) \leq e^{\|R^{-1}\| \varepsilon^{2\nu}} \mathbb{1}_{\{|Rx| \leq \varepsilon^\nu\}} - (|Rx|^2 - \varepsilon^{2\nu}) e^{\frac{\langle x, Rx \rangle}{\varepsilon^{2\nu}}} \mathbb{1}_{\{|Rx| \geq \varepsilon^\nu\}} \quad \forall x \in \mathbb{R}^d, \quad (3.26)$$

for all  $\varepsilon < \varepsilon_0$ . Note that (3.26) is a Foster–Lyapunov equation and  $F$  is inf-compact. Therefore  $\check{v}^\varepsilon$  is a stable Markov control with invariant measure  $\check{\mu}^\varepsilon$ . Thus, integrating (3.26) with respect to the invariant probability measure  $\check{\mu}^\varepsilon$ , we obtain

$$\int_{\{|Rx| \geq \varepsilon^\nu\}} (|Rx|^2 - \varepsilon^{2\nu}) e^{\frac{\langle x, Rx \rangle}{\varepsilon^{2\nu}}} \check{\mu}^\varepsilon(dx) \leq e^{\|R^{-1}\| \varepsilon^{2\nu}} \quad \forall \varepsilon < \varepsilon_0. \quad (3.27)$$

For any  $a \in (0, 1)$  we have

$$|y|^2 \leq \frac{|y|^2 - a^4}{1 - a^2} \quad \text{if } |y| \geq a. \quad (3.28)$$

Thus using (3.27), and applying (3.28) with  $a = \varepsilon^{\nu/2}$ , and the inequality  $\langle x, Rx \rangle \geq \|R\|^{-1}|Rx|^2$ , we obtain

$$\begin{aligned} \int_{|Rx| \geq \varepsilon^{\nu/2}} |Rx|^2 \check{\mu}^\varepsilon(dx) &\leq \int_{|Rx| \geq \varepsilon^{\nu/2}} \frac{|Rx|^2 - \varepsilon^{2\nu}}{1 - \varepsilon^\nu} e^{-\|R\|^{-1}\varepsilon^{-\nu}} e^{\frac{\langle x, Rx \rangle}{\varepsilon^{2\nu}}} \check{\mu}^\varepsilon(dx) \\ &\leq \frac{1}{1 - \varepsilon^\nu} e^{-\|R\|^{-1}\varepsilon^{-\nu}} \int_{|Rx| \geq \varepsilon^\nu} (|Rx|^2 - \varepsilon^{2\nu}) e^{\frac{\langle x, Rx \rangle}{\varepsilon^{2\nu}}} \check{\mu}^\varepsilon(dx) \\ &\leq e^{\|R^{-1}\|} \frac{\varepsilon^{2\nu}}{1 - \varepsilon^\nu} e^{-\|R\|^{-1}\varepsilon^{-\nu}} \quad \forall \varepsilon < \varepsilon_0. \end{aligned} \quad (3.29)$$

Similarly, by (3.27), and using the inequality  $(N^2 - 1)|y|^2 \leq N^2(|y|^2 - \varepsilon^{2\nu})$  if  $|y| \geq N\varepsilon^\nu$  for any  $N \geq 2$ , we obtain

$$\int_{|Rx| \geq N\varepsilon^\nu} |Rx|^2 \check{\mu}^\varepsilon(dx) \leq e^{\|R^{-1}\|} \frac{N^2 \varepsilon^{2\nu}}{N^2 - 1} e^{-N^{-2}\varepsilon^{-2\nu}\|R\|^{-1}} \quad \forall \varepsilon < \varepsilon_0. \quad (3.30)$$

Also, since by definition  $\check{v}^\varepsilon = 0$  for  $|Rx| \leq \varepsilon^{\nu/2}$ , and  $|\check{v}^\varepsilon(x)| \leq \tilde{C}_m \frac{|x|}{\varepsilon}$  by (2.2), it follows by (3.29) that

$$\int_{\mathbb{R}^d} |\check{v}^\varepsilon(x)|^2 \check{\mu}^\varepsilon(dx) \leq \|R^{-1}\|^2 \frac{\tilde{C}_m^2}{1 - \varepsilon^\nu} e^{\|R^{-1}\| \varepsilon^{2\nu-2}} e^{-\|R\|^{-1}\varepsilon^{-\nu}} \quad \forall \varepsilon < \varepsilon_0. \quad (3.31)$$

Then (3.20) follows from (3.29) and (3.31), by choosing a common constant  $c_0$ .

Consider the ‘scaled’ diffusion

$$d\hat{X}_t = \hat{b}^\varepsilon(\hat{X}_t) dt + dW_t, \quad t \geq 0,$$

where

$$\hat{b}^\varepsilon := \frac{m(\varepsilon^\nu x) + \varepsilon \check{v}^\varepsilon(\varepsilon^\nu x)}{\varepsilon^\nu}.$$

and let  $\hat{\mu}^\varepsilon$  denote its invariant probability measure. It  $\check{\varrho}^\varepsilon$  and  $\hat{\varrho}^\varepsilon$  denote the densities of  $\check{\mu}^\varepsilon$  and  $\hat{\mu}^\varepsilon$  respectively, then  $\varepsilon^{\nu d} \check{\varrho}^\varepsilon(\varepsilon^\nu x) = \hat{\varrho}^\varepsilon(x)$  for all  $x \in \mathbb{R}^d$ . Substituting  $x = \varepsilon^\nu y$  in (3.27) we deduce that the family of probability measures  $\{\hat{\mu}^\varepsilon : \varepsilon \in (0, 1)\}$  is tight. The (discontinuous) drift  $\hat{b}^\varepsilon$  converges to  $Mx$  as  $\varepsilon \searrow 0$ , uniformly on compact sets. This implies that  $\hat{\varrho}^\varepsilon$  converges, as  $\varepsilon \searrow 0$ , to the Gaussian density  $\rho_\Sigma$  with mean 0 and covariance matrix  $\Sigma$ , given by  $M\Sigma + \Sigma M^\top = -I$ , i.e.,  $\Sigma = \frac{1}{4}R^{-1}$ , uniformly on compact sets. Indeed, since  $\hat{b}^\varepsilon$  is locally bounded uniformly in  $\varepsilon \in (0, 1)$ , and the family  $\{\hat{\mu}^\varepsilon, \varepsilon \in (0, 1)\}$  is tight, the densities  $\hat{\varrho}^\varepsilon$  of  $\hat{\mu}^\varepsilon$  are locally Hölder equicontinuous (see Lemma 3.2.4 in [1]). Let  $\hat{\varrho}$  be any limit point of  $\hat{\varrho}^\varepsilon$  along some sequence  $\varepsilon_n \searrow 0$ . Since  $\{\hat{\mu}^\varepsilon : \varepsilon \in (0, 1)\}$  is tight it follows that  $\hat{\varrho}^{\varepsilon_n}$  also converges in  $L^1(\mathbb{R}^d)$ , as  $n \rightarrow \infty$ , and hence  $\int_{\mathbb{R}^d} \hat{\varrho}(x) dx = 1$ . With  $\hat{\mathcal{L}}^\varepsilon := \frac{1}{2}\Delta + \langle \hat{b}^\varepsilon, \nabla \rangle$  and  $\hat{\mathcal{L}}^0 := \frac{1}{2}\Delta + \langle Mx, \nabla \rangle$ , and since  $\int_{\mathbb{R}^d} \hat{\mathcal{L}}^\varepsilon f(x) \hat{\varrho}^\varepsilon(x) dx = 0$  for all  $f \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ , we have

$$\int_{\mathbb{R}^d} \hat{\mathcal{L}}^0 f(x) \hat{\varrho}(x) dx = \int_{\mathbb{R}^d} (\hat{\mathcal{L}}^0 f(x) - \hat{\mathcal{L}}^\varepsilon f(x)) \hat{\varrho}(x) dx + \int_{\mathbb{R}^d} \hat{\mathcal{L}}^\varepsilon f(x) (\hat{\varrho}(x) - \hat{\varrho}^\varepsilon(x)) dx \quad (3.32)$$

for all  $f \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ . It is clear that both terms on the right hand side of (3.32) converge to 0 as  $\varepsilon = \varepsilon_n \searrow 0$ . This implies that  $\hat{\varrho}$  is the density of the invariant probability measure of the diffusion  $dX_t = MX_t dt + dW_t$ , which is Gaussian as claimed.

Since the Gaussian density  $\rho_\Sigma$  has zero mean, then by uniform integrability implied by (3.30) we have

$$\varepsilon^{-\nu} \int_{\mathbb{R}^d} (D\ell(0)x) \check{\mu}^\varepsilon(dx) \xrightarrow[\varepsilon \searrow 0]{} 0. \quad (3.33)$$

It follows by (3.30) that for some constant  $\kappa_1 > 0$  we have  $\int_{\mathbb{R}^d} |x|^2 \check{\mu}^\varepsilon(dx) < \kappa_1$  for all  $\varepsilon < \varepsilon_0$ . Thus, using (3.15), we obtain

$$\varepsilon^{-\nu} \int_{\mathbb{R}^d} |\ell(x) - \ell(0) - D\ell(0)x| \check{\mu}^\varepsilon(dx) \leq \kappa_1 \tilde{C}_\ell \varepsilon^\nu. \quad (3.34)$$

Combining (3.33)–(3.34), we obtain (3.21).

Next we turn to part (ii). Consider the control  $v^\varepsilon(x) = \varepsilon^{-1}(Mx - m(x))$  for  $x \in \mathbb{R}^d$ . Then  $m(x) + \varepsilon v^\varepsilon(x) = Mx$  and the associated invariant measure  $\mu^\varepsilon$  is Gaussian with mean 0 and covariance matrix  $\varepsilon^{2\nu} \Sigma$ . Using the bound in (2.2), we obtain

$$\int_{\mathbb{R}^d} |v^\varepsilon|^2 d\mu^\varepsilon \leq \int_{\mathbb{R}^d} \tilde{C}_m^2 \varepsilon^{-2} |x|^4 \mu^\varepsilon(dx) \in \mathcal{O}(\varepsilon^{4\nu-2}).$$

Since  $\mu^\varepsilon$  has zero mean, using a triangle inequality, and (3.15), as in the proof of Lemma 3.2, we obtain

$$\left| \int_{\mathbb{R}^d} (\ell(x) - \ell(0)) \mu^\varepsilon(dx) \right| \leq \varepsilon^{2\nu} \tilde{C}_\ell \text{trace}(\Sigma).$$

Since  $4\nu - 2 < 2\nu$  for  $\nu < 1$ , we obtain that  $\beta_*^\varepsilon \leq \mathfrak{J}_s + \mathcal{O}(\varepsilon^{4\nu-2})$ . On the other hand, by part (1) we already know that  $\beta_*^\varepsilon \leq \mathfrak{J}_s + \mathfrak{o}(\varepsilon^\nu)$ . To complete the proof we observe that  $\nu \leq 4\nu - 2$  for  $\nu \geq 2/3$  and  $\nu > 4\nu - 2$  for  $\nu < 2/3$ .  $\square$

**3.3. Results concerning the subcritical regime.** By Lemma 3.5 we can always find a stable admissible control such that the corresponding invariant probability measure concentrates on a stable equilibrium point as  $\varepsilon \searrow 0$ , while keeping the ergodic cost in (1.3) bounded, uniformly in  $\varepsilon \in (0, 1)$ . Now we proceed to show that for  $\nu < 1$ ,  $\eta_*^\varepsilon$  concentrates on  $\mathcal{S}_s$ .

**Lemma 3.6.** *Suppose  $\nu < 1$ . Then  $\eta_*^\varepsilon(\mathcal{S} \setminus \mathcal{S}_s) \xrightarrow[\varepsilon \searrow 0]{} 0$ , and  $\lim_{\varepsilon \searrow 0} \beta_*^\varepsilon = \mathfrak{J}_s$ .*

*Proof.* We argue by contradiction. Suppose that  $\limsup_{\varepsilon \searrow 0} \eta_*^\varepsilon(B_r(z)) > 0$  for some  $r > 0$  and  $z \notin \mathcal{S}_s$ .

In Theorem 2.2 we may select  $a_z$  such that  $a_z \neq a'_z$  for  $z \neq z'$ . Thus by Theorem 2.2 (ii), there exists  $\delta > 0$  be such that the interval  $(\mathcal{V}(z) - 3\delta, \mathcal{V}(z) + 3\delta)$  contains no other critical values of  $\mathcal{V}$  other than  $\mathcal{V}(z)$ . Let  $\varphi \in \mathcal{C}^2(\mathbb{R})$  be such that

- (a)  $\varphi(\mathcal{V}(z) + y) = y$  for  $y \in (\mathcal{V}(z) - \delta, \mathcal{V}(z) + \delta)$ ;
- (b)  $\varphi' \in [0, 1]$  on  $(\mathcal{V}(z) - 2\delta, \mathcal{V}(z) + 2\delta)$ ;
- (c)  $\varphi' = 0$  on  $(\mathcal{V}(z) - 2\delta, \mathcal{V}(z) + 2\delta)^c$ .

Select  $r > 0$  such that

$$\sup_{x \in B_r(z)} |\Delta\mathcal{V}(x) - \Delta\mathcal{V}(z)| < \frac{1}{2} |\Delta\mathcal{V}(z)|. \quad (3.35)$$

Note that by Theorem 2.2 and Lemma 2.3 the function  $\mathcal{V}$  takes distinct values on  $\mathcal{S}$ . Therefore we may also choose this  $r$  small enough so that

$$B_r(z) \subset \{x : |\mathcal{V}(x) - \mathcal{V}(z)| \leq \delta\} \subset B_r^c(S \setminus \{z\}).$$

By the infinitesimal characterization of an invariant probability measure we have

$$\int_{\mathbb{R}^d} \mathcal{L}_{v_*^\varepsilon}^\varepsilon(\varphi \circ \mathcal{V})(x) \eta_*^\varepsilon(dx) = 0,$$

which we write as

$$\begin{aligned} \frac{\varepsilon^{2\nu}}{2} \left( \int_{\mathbb{R}^d} \varphi'(\mathcal{V}) \Delta\mathcal{V} d\eta_*^\varepsilon + \int_{\mathbb{R}^d} \varphi''(\mathcal{V}) |\nabla\mathcal{V}|^2 d\eta_*^\varepsilon \right) + \varepsilon \int_{\mathbb{R}^d} \varphi'(\mathcal{V}) \langle v_*^\varepsilon, \nabla\mathcal{V} \rangle d\eta_*^\varepsilon \\ + \int_{\mathbb{R}^d} \varphi'(\mathcal{V}) \langle m, \nabla\mathcal{V} \rangle d\eta_*^\varepsilon = 0. \end{aligned} \quad (3.36)$$

Recall the definition of the optimal control effort  $\mathcal{G}_*^\varepsilon$  in (1.17), and also define

$$\begin{aligned} \zeta^\varepsilon &:= \left( \int_{\mathbb{R}^d} \varphi'(\mathcal{V}) |\nabla\mathcal{V}|^2 d\eta_*^\varepsilon \right)^{1/2}, \\ \xi_1^\varepsilon &:= \frac{1}{2} \int_{\mathbb{R}^d} \varphi'(\mathcal{V}) \Delta\mathcal{V} d\eta_*^\varepsilon, \quad \xi_2^\varepsilon := \frac{1}{2} \int_{\mathbb{R}^d} \varphi''(\mathcal{V}) |\nabla\mathcal{V}|^2 d\eta_*^\varepsilon, \end{aligned} \quad (3.37)$$

and  $\xi^\varepsilon := \xi_1^\varepsilon + \xi_2^\varepsilon$ . By the Cauchy–Schwarz inequality we have

$$\left| \int_{\mathbb{R}^d} \varphi'(\mathcal{V}) \langle v_*^\varepsilon, \nabla\mathcal{V} \rangle d\eta_*^\varepsilon \right| \leq \|\sqrt{\varphi'}\|_\infty \sqrt{2\mathcal{G}_*^\varepsilon} \zeta^\varepsilon \leq \sqrt{2\mathcal{G}_*^\varepsilon} \zeta^\varepsilon. \quad (3.38)$$

By Theorem 2.2 (iv) we have  $C_0 (\zeta^\varepsilon)^2 \leq - \int_{\mathbb{R}^d} \varphi'(\mathcal{V}) \langle m, \nabla\mathcal{V} \rangle d\eta_*^\varepsilon$ . Therefore, by (3.36) and (3.38) we obtain

$$C_0 (\zeta^\varepsilon)^2 - \varepsilon \sqrt{2\mathcal{G}_*^\varepsilon} \zeta^\varepsilon - \varepsilon^{2\nu} \xi^\varepsilon \leq 0. \quad (3.39)$$

We write

$$\xi_1^\varepsilon = \int_{B_r(z)} \varphi'(\mathcal{V}) \Delta\mathcal{V} d\eta_*^\varepsilon + \int_{B_r^c(z)} \varphi'(\mathcal{V}) \Delta\mathcal{V} d\eta_*^\varepsilon. \quad (3.40)$$

Since  $\mathcal{V}$  is inf-compact, it follows that  $\varphi \circ \mathcal{V}$  is constant outside a compact set. Therefore, the support of  $\varphi'(\mathcal{V}(\cdot))$  is compact, and as a result  $\Delta\mathcal{V}$  is bounded on this set. By (3.35), (3.40), Theorem 2.2 (iii), and since  $\eta_*^\varepsilon(B_r^c(\mathcal{S})) \searrow 0$  as  $\varepsilon \searrow 0$  (by Lemma 3.1), we obtain

$$\limsup_{\varepsilon \searrow 0} (-\xi_1^\varepsilon) \geq -\frac{1}{2} \Delta\mathcal{V}(z) \limsup_{\varepsilon \searrow 0} \eta_*^\varepsilon(B_r(z)) > 0. \quad (3.41)$$

On the other hand, since  $\varphi''(\mathcal{V}) = 0$  on some open neighborhood of  $\mathcal{S}$ , it follows that  $\xi_2^\varepsilon \rightarrow 0$  as  $\varepsilon \searrow 0$ . Therefore, we have  $\limsup_{\varepsilon \searrow 0} (-\xi^\varepsilon) > 0$ . However, since the discriminant of (3.39) must be nonnegative, we obtain

$$\varepsilon^2 \mathcal{G}_*^\varepsilon \geq -2 C_0 \varepsilon^{2\nu} \xi^\varepsilon, \quad (3.42)$$

which leads to a contradiction. Hence,  $\eta_*^\varepsilon(\mathcal{S} \setminus \mathcal{S}_s) \xrightarrow[\varepsilon \searrow 0]{} 0$ . This implies that  $\liminf_{\varepsilon \searrow 0} \beta_*^\varepsilon \geq \mathfrak{J}_s$ , which combined with Lemma 3.5 (ii), results in equality for the limit as claimed.  $\square$

We revisit the subcritical regime in Corollary 4.2 to obtain a lower bound for  $\beta_*^\varepsilon$ .

It is worthwhile at this point to present the following one-dimensional example, which shows how the value of  $\beta_*^\varepsilon$  for small  $\varepsilon$  bifurcates as we cross the critical regime.

**Example 3.7.** Let  $d = 1$ ,  $m(x) = Mx$ , and  $\ell(x) = \frac{1}{2}Lx^2$ , with  $M > 0$  and  $L > 0$ . Then the solution to (1.14) is:

$$\begin{aligned} V^\varepsilon &= \frac{M + \sqrt{M^2 + L\varepsilon^2}}{2\varepsilon^2} x^2, \\ \beta_*^\varepsilon &= \frac{\varepsilon^{2\nu-2}}{2} (M + \sqrt{M^2 + L\varepsilon^2}). \end{aligned}$$

Note that  $\beta_*^\varepsilon \rightarrow \ell(0) = 0$ ,  $\beta_*^\varepsilon \rightarrow M$ , and  $\beta_*^\varepsilon \rightarrow \infty$ , as  $\varepsilon \searrow 0$ , when  $\nu > 1$ ,  $\nu = 1$ , and  $\nu < 1$ , respectively.

#### 4. CONCENTRATION BOUNDS FOR THE OPTIMAL STATIONARY DISTRIBUTION

We start with the following lemma, which is valid for all  $\nu$ .

**Lemma 4.1.** *For any bounded domain  $G$  there exists a constant  $\hat{\kappa}_0 = \hat{\kappa}_0(G, \nu)$  such that*

$$\int_G (\text{dist}(x, \mathcal{S}))^2 \eta_*^\varepsilon(dx) \leq \hat{\kappa}_0 \varepsilon^{2(\nu \wedge 2)} \quad \forall \nu > 0, \quad \forall \varepsilon \in (0, 1), \quad (4.1)$$

where  $\text{dist}(x, \mathcal{S})$  denotes the Euclidean distance of  $x$  from the set  $\mathcal{S}$ .

*Proof.* We fix some bounded domain  $G$  which, without loss of generality contains  $\mathcal{S}$ , and choose some number  $\delta$  such that  $\delta \geq \sup_{x \in G} \mathcal{V}(x)$ . Without loss of generality assume that  $\ell(x) > \mathfrak{J}$  for all  $x \in G^c$ , otherwise we enlarge  $G$ . Let  $\tilde{\varphi}: \mathbb{R} \rightarrow \mathbb{R}$  be a smooth function such that

- (a)  $\tilde{\varphi}(y) = y$  for  $y \in (-\infty, \delta)$ ;
- (b)  $\tilde{\varphi}' \in (0, 1)$  on  $(\delta, 2\delta)$ ;
- (c)  $\tilde{\varphi}' = 0$  on  $[2\delta, \infty)$ ;
- (d)  $\tilde{\varphi}'' \leq 0$ .

Define  $\tilde{\zeta}^\varepsilon$ ,  $\tilde{\xi}_1^\varepsilon$ , and  $\tilde{\xi}_2^\varepsilon$ , as in (3.37) by replacing  $\varphi$  with  $\tilde{\varphi}$ , and let  $\tilde{\xi}^\varepsilon := \tilde{\xi}_1^\varepsilon + \tilde{\xi}_2^\varepsilon$ . As in (3.39) we obtain

$$C_0 (\tilde{\zeta}^\varepsilon)^2 - \varepsilon \sqrt{2G_*^\varepsilon} \tilde{\zeta}^\varepsilon - \varepsilon^{2\nu} \tilde{\xi}^\varepsilon \leq 0. \quad (4.2)$$

By Theorem 2.2 (iv) we have

$$\int_{\{x : \mathcal{V}(x) \leq \delta\}} (\text{dist}(x, \mathcal{S}))^2 \eta_*^\varepsilon(dx) \leq C_0^{-1} (\tilde{\zeta}^\varepsilon)^2. \quad (4.3)$$

By an application of Young's inequality to (4.2), we obtain  $\frac{C_0}{2} (\tilde{\zeta}^\varepsilon)^2 - \frac{1}{C_0} \varepsilon^2 G_*^\varepsilon - \varepsilon^{2\nu} \tilde{\xi}^\varepsilon \leq 0$  and hence  $\tilde{\zeta}^\varepsilon \in \mathcal{O}(\varepsilon^{\nu \wedge 2})$ . Thus (4.1) follows by (4.3).  $\square$

**Corollary 4.2.** *Suppose  $\nu \geq 1$ . Then following hold.*

- (a) *The optimal control effort  $\mathfrak{G}_*^\varepsilon$  satisfies*

$$\begin{aligned} \mathfrak{G}_*^\varepsilon &\in \mathcal{O}(\varepsilon^{\nu \wedge 2}) && \text{if } \mathfrak{J} = \mathfrak{J}_s \text{ and } \nu > 1, \text{ or if } \mathfrak{J}_c = \mathfrak{J}_s \text{ and } \nu = 1, \\ \mathfrak{G}_*^\varepsilon &\in \mathcal{O}(\varepsilon^{(2\nu-2)\wedge 2}) && \text{if } \mathfrak{J} < \mathfrak{J}_s \text{ and } \nu > 1, \end{aligned} \quad (4.4)$$

and

$$\liminf_{\varepsilon \searrow 0} \frac{1}{\varepsilon^{2\nu-2}} \mathfrak{G}_*^\varepsilon > 0 \quad \text{if } \mathfrak{J} < \mathfrak{J}_s, \text{ and } \nu > 1. \quad (4.5)$$

(b)  $\beta_*^\varepsilon - \mathfrak{J} \geq \mathcal{O}(\varepsilon^{\nu \wedge 2})$  for  $\nu > 1$ .

*Proof.* Select a domain  $G$  as in the proof of Lemma 4.1. Define  $\tilde{\zeta}^\varepsilon$ ,  $\tilde{\xi}_1^\varepsilon$ , and  $\tilde{\xi}_2^\varepsilon$  as in (3.37) by replacing  $\varphi$  with  $\tilde{\varphi}$ , and let  $\tilde{\xi}^\varepsilon := \tilde{\xi}_1^\varepsilon + \tilde{\xi}_2^\varepsilon$ . Then (4.2) holds, and thus  $\tilde{\zeta}^\varepsilon \in \mathcal{O}(\varepsilon^{\nu \wedge 2})$ . Recall the notation in Definition 1.10. With  $C_\ell$  a Lipschitz constant for  $\ell$ , and some fixed  $\bar{z} \in \mathcal{Z}$ , we have

$$\ell(x) - \mathfrak{J} = (\ell(x) - \ell(z)) + (\ell(z) - \ell(\bar{z})) \geq -C_\ell|x - z| \quad \forall z \in \mathcal{S}, \quad \forall x \in \mathbb{R}^d,$$

since  $\ell(z) - \ell(\bar{z}) \geq 0$  for all  $z \in \mathcal{S}$ . Therefore, we obtain

$$\ell(x) - \mathfrak{J} \geq -C_\ell \text{dist}(x, \mathcal{S}) \quad \forall x \in \mathbb{R}^d, \quad (4.6)$$

and using the Cauchy–Schwarz inequality, and the assumption that  $\ell(x) > \mathfrak{J}$  on  $G^c$ , we deduce from (4.6) and Theorem 2.2 (iv) that

$$\int_{\mathbb{R}^d} \ell d\eta_*^\varepsilon - \mathfrak{J} \geq \int_G (\ell(x) - \mathfrak{J}) d\eta_*^\varepsilon \geq -\frac{C_\ell}{\sqrt{C_0}} \tilde{\zeta}^\varepsilon. \quad (4.7)$$

Thus by (4.7) and non-negativity of  $\mathcal{G}^\varepsilon$  we have

$$-\frac{C_\ell}{\sqrt{C_0}} \tilde{\zeta}^\varepsilon \leq \beta_*^\varepsilon - \mathfrak{J}. \quad (4.8)$$

By (4.7)–(4.8), we obtain

$$\begin{aligned} \mathcal{G}_*^\varepsilon &\leq \beta_*^\varepsilon - \int_{\mathbb{R}^d} \ell d\eta_*^\varepsilon \\ &\leq \beta_*^\varepsilon - \mathfrak{J} + \mathfrak{J} - \int_{\mathbb{R}^d} \ell d\eta_*^\varepsilon \\ &\leq \beta_*^\varepsilon - \mathfrak{J} + \frac{C_\ell}{\sqrt{C_0}} \tilde{\zeta}^\varepsilon. \end{aligned} \quad (4.9)$$

By an application of Young's inequality to (4.2), we obtain  $\frac{C_0}{2} (\tilde{\zeta}^\varepsilon)^2 - \frac{1}{C_0} \varepsilon^2 \mathcal{G}_*^\varepsilon - \varepsilon^{2\nu} \tilde{\xi}^\varepsilon \leq 0$ , and thus

$$\tilde{\zeta}^\varepsilon \leq \frac{\sqrt{2}}{C_0} \varepsilon \sqrt{\mathcal{G}_*^\varepsilon} + \frac{\sqrt{2}}{\sqrt{C_0}} \varepsilon^\nu \sqrt{|\tilde{\xi}^\varepsilon|}. \quad (4.10)$$

Combining (4.9)–(4.10), and using again Young's inequality, we have

$$\frac{1}{2} \mathcal{G}_*^\varepsilon \leq \beta_*^\varepsilon - \mathfrak{J} + \frac{C_\ell}{C_0^3} \varepsilon^2 + \frac{\sqrt{2} C_\ell}{C_0} \varepsilon^\nu \sqrt{|\tilde{\xi}^\varepsilon|}. \quad (4.11)$$

By Lemma 3.3 and (4.11) we obtain  $\mathcal{G}_*^\varepsilon \in \mathcal{O}(\varepsilon^{\nu \wedge 2})$  if  $\mathfrak{J} = \mathfrak{J}_s$  for  $\nu > 1$ , or if  $\mathfrak{J}_c = \mathfrak{J}_s$  and  $\nu = 1$ . We also obtain  $\mathcal{G}_*^\varepsilon \in \mathcal{O}(\varepsilon^{2 \wedge (2\nu-2)})$  if  $\mathfrak{J} < \mathfrak{J}_s$  and  $\nu > 1$ . Thus (4.4) holds.

If  $\mathfrak{J} < \mathfrak{J}_s$  and  $\nu > 1$ , then  $\mathcal{Z} \subset \mathcal{S} \setminus \mathcal{S}_s$ , and  $\mathfrak{S} \subset \mathcal{Z}$  by Lemma 3.3. Fix some  $z \in \mathfrak{S}$ . Then  $\liminf_{\varepsilon \searrow 0} \eta_*^\varepsilon(B_r(z)) > 0$  for any  $r > 0$ . Also  $\Delta \mathcal{V}(z) < 0$  by Theorem 2.2. Therefore (3.41) holds, with ‘lim inf’ replacing the ‘lim sup’. Expanding  $\tilde{\xi}_1^\varepsilon$  as in (3.40), and arguing as in Lemma 3.6 it follows that (3.41) with ‘lim inf’ also holds for  $\tilde{\xi}^\varepsilon$ . In fact, it easily follows that for some constant  $\kappa_1$ , we have

$$\liminf_{\varepsilon \searrow 0} (-\tilde{\xi}^\varepsilon) \geq \min_{z \in \mathcal{Z}} \kappa_1(-\frac{1}{2} \Delta \mathcal{V}(z)). \quad (4.12)$$

The discriminant of the quadratic polynomial in (4.2) is nonnegative and this implies that

$$\varepsilon^2 \mathcal{G}_*^\varepsilon \geq -2 C_0 \varepsilon^{2\nu} \tilde{\xi}^\varepsilon, \quad (4.13)$$

in direct analogy with (3.42). Thus, (4.5) follows by (4.12) and (4.13). This completes the proof of part (a).

Since  $\tilde{\zeta}^\varepsilon \in \mathcal{O}(\varepsilon^{\nu \wedge 2})$ , we obtain  $\beta_*^\varepsilon - \mathfrak{J} \geq \mathcal{O}(\varepsilon^{\nu \wedge 2})$  by (4.8). This proves part (b), and completes the proof.  $\square$

We define the following scaled quantities.

**Definition 4.3.** For  $z \in \mathcal{S}$ , and  $V^\varepsilon$  as in Theorem 1.4, we define

$$\widehat{V}_z^\varepsilon(x) := V^\varepsilon(\varepsilon^\nu x + z), \quad x \in \mathbb{R}^d.$$

and

$$\tilde{V}^\varepsilon := \varepsilon^2 V^\varepsilon, \quad \check{V}_z^\varepsilon := \varepsilon^{2(1-\nu)} \widehat{V}_z^\varepsilon.$$

We also define the ‘scaled’ vector field and penalty by

$$\widehat{m}_z^\varepsilon(x) := \frac{m(\varepsilon^\nu x + z)}{\varepsilon^\nu}, \quad \widehat{\ell}_z^\varepsilon(x) := \ell(\varepsilon^\nu x + z).$$

The next lemma shows provides estimates for the growth of  $\nabla \widehat{V}_z^\varepsilon$ , and  $\nabla \tilde{V}^\varepsilon$ .

**Lemma 4.4.** Assume  $\nu \in (0, 2]$ , and let  $\widehat{V}_z^\varepsilon$ ,  $\tilde{V}^\varepsilon$ , and  $\check{V}_z^\varepsilon$ , be as in Definition 4.3. Then

(a) Under the restriction that  $z \in \mathcal{Z}$  when  $\nu \in (1, 2]$ , there exists a constant  $\check{c}_0$  such that

$$|\nabla \check{V}_z^\varepsilon(x)| \leq \check{c}_0 (1 + |x|) \quad \forall \varepsilon \in (0, 1), \quad \forall x \in \mathbb{R}^d. \quad (4.14)$$

(b) The bound in (4.14) also holds for  $\tilde{V}^\varepsilon$  for all  $\nu \in (0, 2]$ , with no restrictions on  $z$ .

*Proof.* By (1.14), the function  $\check{V}_z^\varepsilon$  satisfies

$$\frac{1}{2} \Delta \check{V}_z^\varepsilon(x) + \langle \widehat{m}_z^\varepsilon(x), \nabla \check{V}_z^\varepsilon(x) \rangle - \frac{1}{2} |\nabla \check{V}_z^\varepsilon(x)|^2 = \varepsilon^{2(1-\nu)} (\beta_*^\varepsilon - \widehat{\ell}_z^\varepsilon(x)). \quad (4.15)$$

Since  $\ell$  is Lipschitz, the gradient of the map  $x \mapsto \varepsilon^{2(1-\nu)} (\widehat{\ell}_z^\varepsilon(x) - \ell(z))$  is bounded in  $\mathbb{R}^d$ , uniformly in  $\varepsilon \in (0, 1)$ , and  $\nu \in (0, 2]$ . Similarly,  $|\widehat{m}_z^\varepsilon(x)|$ ,  $\|D\widehat{m}_z^\varepsilon(x)\|$  and  $\|D^2\widehat{m}_z^\varepsilon(x)\|$ , are bounded in  $\mathbb{R}^d$ , uniformly in  $\varepsilon \in (0, 1)$ , and  $\nu \in (0, 2]$ . By Theorem 1.11 (i), which is established in Corollary 4.2, the constants  $\varepsilon^{2(1-\nu)} (\beta_*^\varepsilon - \ell(z))$  are bounded uniformly in  $\varepsilon \in (0, 1)$ , and  $\nu \in (1, 2]$  for  $z \in \mathcal{Z}$ . Applying [29, Lemma 5.1] to (4.15) it follows that  $\check{V}_z^\varepsilon$  satisfies (4.14) if  $\nu \in (1, 2]$  and  $z \in \mathcal{Z}$ . On the other hand, if  $\nu \in (0, 1]$ , then the gradient of the right hand side of (4.15) is bounded in  $\mathbb{R}^d$ , uniformly in  $\varepsilon \in (0, 1)$ , and the restriction  $z \in \mathcal{Z}$  is not needed. This completes the proof of part (a).

Next show that (4.14) holds for  $\tilde{V}^\varepsilon$ . Fix an arbitrary  $z \in \mathcal{Z}$ . We have

$$\begin{aligned} \nabla_x V^\varepsilon(x + z) &= \varepsilon^{-\nu} \nabla_y \widehat{V}_z^\varepsilon(y) \Big|_{y=\varepsilon^{-\nu}x} \\ &\leq \frac{\varepsilon^{-\nu}}{\varepsilon^{2(1-\nu)}} \check{c}_0 (1 + |\varepsilon^{-\nu}x|) \\ &= \frac{\check{c}_0}{\varepsilon^2} (\varepsilon^\nu + |x|), \end{aligned}$$

where in the inequality we use the identity  $\widehat{V}_z^\varepsilon = \varepsilon^{2(\nu-1)} \check{V}_z^\varepsilon$  and (4.14). Since  $\tilde{V}^\varepsilon = \varepsilon^2 V^\varepsilon$ , this proves the property for  $\tilde{V}^\varepsilon$ . This completes the proof.  $\square$

We continue with a version of Lemma 4.1 for unbounded domains.

**Proposition 4.5.** Let  $\nu \in (0, 2]$ . Then for any  $k \in \mathbb{N}$  and  $r > 0$ , there exist constants and  $\hat{\kappa}_1 = \hat{\kappa}_1(k)$  and  $\hat{\kappa}_2 = \hat{\kappa}_2(k)$  such that with  $\hat{r}(\varepsilon) := \hat{\kappa}_2 \varepsilon^{\nu \wedge 1}$  we have

$$\int_{B_{\hat{r}(\varepsilon)}^c(\mathcal{S})} (\text{dist}(x, \mathcal{S}))^{2k} \eta_*^\varepsilon(dx) \leq \hat{\kappa}_1 \varepsilon^{2(\nu \wedge 1)} \quad \forall \varepsilon \in (0, 1).$$

*Proof.* Let  $\tilde{V}^\varepsilon := \varepsilon^2 V^\varepsilon$ . Since  $V^\varepsilon(0) = 0$ , by Lemma 4.4 the function  $\tilde{V}^\varepsilon = \varepsilon^2 V^\varepsilon$  is locally bounded, uniformly in  $\varepsilon > 0$ . Applying the operator

$$\mathcal{L}_*^\varepsilon := \frac{\varepsilon^{2\nu}}{2} \Delta + \langle m - \varepsilon^2 \nabla V^\varepsilon, \nabla \rangle$$

to the function  $\mathcal{V}^{2k} e^{\tilde{V}^\varepsilon}$  and using the identity  $\mathcal{L}_*^\varepsilon[\tilde{V}^\varepsilon] = \varepsilon^2(\beta_*^\varepsilon - \ell) - \frac{1}{2} |\nabla \tilde{V}^\varepsilon|^2$ , and rearranging terms we obtain

$$\begin{aligned} \mathcal{L}_*^\varepsilon[\mathcal{V}^{2k} e^{\tilde{V}^\varepsilon}] &= \mathcal{V}^{2k} \mathcal{L}_*^\varepsilon[e^{\tilde{V}^\varepsilon}] + e^{\tilde{V}^\varepsilon} \mathcal{L}_*^\varepsilon[\mathcal{V}^{2k}] + 2k \varepsilon^{2\nu} \mathcal{V}^{(2k-1)} e^{\tilde{V}^\varepsilon} \nabla \tilde{V}^\varepsilon \nabla \mathcal{V} \\ &= \mathcal{V}^{2k} e^{\tilde{V}^\varepsilon} \left[ \varepsilon^2(\beta_*^\varepsilon - \ell) + k \varepsilon^{2\nu} \frac{\Delta \mathcal{V}}{\mathcal{V}} - \frac{1 - \varepsilon^{2\nu}}{2} \left( \nabla \tilde{V}^\varepsilon + 2k \frac{\nabla \mathcal{V}}{\mathcal{V}} \right)^2 \right. \\ &\quad \left. + 2k \frac{\langle m, \nabla \mathcal{V} \rangle}{\mathcal{V}} + k(2k - \varepsilon^{2\nu}) \frac{|\nabla \mathcal{V}|^2}{\mathcal{V}^2} \right]. \end{aligned} \quad (4.16)$$

By (2.7), and since  $\bar{\mathcal{V}}$  has strict quadratic growth and  $\nabla \bar{\mathcal{V}}$  is Lipschitz by Hypothesis 1.1, and  $\mathcal{V}$  agrees with  $\bar{\mathcal{V}}$  outside a compact set, it follows that  $\frac{|\nabla \mathcal{V}|^2}{\mathcal{V}}$  is bounded on  $\mathbb{R}^d$ . Therefore, in view of the bounds in (2.1) and (2.7), we can add a positive constant to  $\mathcal{V}$  so that

$$2 \frac{\langle m, \nabla \mathcal{V} \rangle}{\mathcal{V}} + (2k - \varepsilon^{2\nu}) \frac{|\nabla \mathcal{V}|^2}{\mathcal{V}^2} \leq \frac{\langle m, \nabla \mathcal{V} \rangle}{\mathcal{V}} \quad \text{on } \mathbb{R}^d, \quad \forall \varepsilon > 0. \quad (4.17)$$

The constant is selected so that  $\mathcal{V} \geq 1$  on  $\mathbb{R}^d$ . Define

$$G_0^\varepsilon := \varepsilon^{2-2\wedge 2\nu} (\beta_*^\varepsilon - \ell) - \frac{1 - \varepsilon^{2\nu}}{2\varepsilon^{2\wedge 2\nu}} \left| \nabla \tilde{V}^\varepsilon + 2k \frac{\nabla \mathcal{V}}{\mathcal{V}} \right|^2.$$

Since  $\ell$  is inf-compact, there exists  $r_0 > 0$  such that  $G_0^\varepsilon \leq 0$  on  $B_{r_0}^c$ . We may choose  $r_0$  large enough so that  $\mathcal{S} \subset B_{r_0}$ . Let  $\kappa_0$  be a bound of  $\beta_*^\varepsilon - \ell$  on  $B_{r_0}$ . Using this bound and (4.16)–(4.17), we obtain

$$\frac{1}{\varepsilon^{2\wedge 2\nu}} \mathcal{L}_*^\varepsilon[\mathcal{V}^{2k} e^{\tilde{V}^\varepsilon}](x) \leq \mathcal{V}^{2k}(x) e^{\tilde{V}^\varepsilon(x)} \left[ \kappa_0 \mathbf{1}_{B_{r_0}}(x) + \frac{k}{\varepsilon^{2\wedge 2\nu}} \frac{\varepsilon^{2\nu} \Delta \mathcal{V}(x) + \langle m(x), \nabla \mathcal{V}(x) \rangle}{\mathcal{V}(x)} \right] \quad (4.18)$$

for all  $x \in \mathbb{R}^d$ , and all  $\varepsilon \in (0, 1)$ . By (2.1) we have

$$\varepsilon^{2\nu} \Delta \mathcal{V}(x) + \langle m(x), \nabla \mathcal{V}(x) \rangle \leq \frac{1}{2} \langle m(x), \nabla \mathcal{V}(x) \rangle \quad (4.19)$$

for all  $x \in \mathbb{R}^d$  such that  $\text{dist}(x, \mathcal{S}) \geq \kappa_1 \varepsilon^\nu$ , with  $\kappa_1 := \sqrt{2C_0^{-1} \|\Delta \mathcal{V}\|_\infty}$ . Using (2.1) once more, if we define  $\kappa_2 := (4k^{-1} C_0^{-1} \kappa_0 \sup_{B_{r_0}} \mathcal{V})^{1/2}$ , then we have

$$\varepsilon^{2\wedge 2\nu} \kappa_0 + \frac{k \langle m(x), \nabla \mathcal{V}(x) \rangle}{4\mathcal{V}(x)} \leq 0 \quad \text{in } \{x \in B_{r_0} : \text{dist}(x, \mathcal{S}) \geq \kappa_2 \varepsilon^{1\wedge \nu}\}. \quad (4.20)$$

Combining (4.18), (4.19), and (4.20), we obtain

$$\frac{1}{\varepsilon^{2\wedge 2\nu}} \mathcal{L}_*^\varepsilon[\mathcal{V}^{2k} e^{\tilde{V}^\varepsilon}](x) \leq \frac{k}{4\varepsilon^{2\wedge 2\nu}} \mathcal{V}^{2k-1}(x) e^{\tilde{V}^\varepsilon(x)} \langle m(x), \nabla \mathcal{V}(x) \rangle \quad (4.21)$$

for all  $x \in \mathbb{R}^d$  such that  $\text{dist}(x, \mathcal{S}) \geq \hat{r}(\varepsilon) := (\kappa_1 \vee \kappa_2) \varepsilon^{\nu \wedge 1}$ . Let  $\kappa_3$  be a bound of the right hand side of (4.18) on  $B_{\hat{r}(\varepsilon)}(\mathcal{S})$ . This bound does not depend on  $\varepsilon$ , since  $\tilde{V}^\varepsilon$  is locally bounded, uniformly in  $\varepsilon \in (0, 1)$ . Then, by (4.18) and (4.21) we obtain

$$\frac{1}{\varepsilon^{2\wedge 2\nu}} \mathcal{L}_*^\varepsilon[\mathcal{V}^{2k} e^{\tilde{V}^\varepsilon}](x) \leq \kappa_3 + \frac{k}{4\varepsilon^{2\wedge 2\nu}} \langle m(x), \nabla \mathcal{V}(x) \rangle \mathcal{V}^{2k-1}(x) e^{\tilde{V}^\varepsilon(x)} \mathbf{1}_{B_{\hat{r}(\varepsilon)}^c(\mathcal{S})}(x) \quad (4.22)$$

for all  $x \in \mathbb{R}^d$ , and  $\varepsilon \in (0, 1)$ .

By the strong maximum principle,  $V^\varepsilon$  attains its infimum in  $\mathbb{R}^d$  in the set  $\{x \in \mathbb{R}^d : \ell(x) \leq \beta_*^\varepsilon\}$ . Therefore,  $\tilde{V}^\varepsilon$  is bounded below in  $\mathbb{R}^d$ , uniformly in  $\varepsilon$ , by Lemma 4.4. Thus, from (4.22) we obtain

$$\int_{B_{\hat{r}(\varepsilon)}^c(\mathcal{S})} \frac{|\langle m(x), \nabla \mathcal{V}(x) \rangle|}{\varepsilon^{2 \wedge 2\nu}} \mathcal{V}^{2k-1}(x) \eta_*^\varepsilon(dx) \leq \frac{4\kappa_3}{k(\inf_{\mathbb{R}^d} e^{\tilde{V}^\varepsilon})} \quad \forall \varepsilon < \varepsilon_0. \quad (4.23)$$

By the strict quadratic growth of  $\mathcal{V}$  mentioned earlier, together with (2.7) and (4.23), there exists a constant  $\kappa_4$ , such that

$$\int_{B_{\hat{r}(\varepsilon)}^c(\mathcal{S})} \frac{1}{\varepsilon^{2 \wedge 2\nu}} (\text{dist}(x, \mathcal{S}))^{4k-1} \eta_*^\varepsilon(dx) \leq \kappa_4 \quad \forall \varepsilon \in (0, 1).$$

This finishes the proof.  $\square$

**Corollary 4.6.** *Let  $D$  be any open set such that  $\mathcal{S}_s \subset D$ . The following hold.*

- (a) *If  $\mathfrak{J} = \mathfrak{J}_s$ , then  $\eta_*^\varepsilon(D^c) \in \mathcal{O}(\varepsilon^{2-\nu})$  for all  $\nu \in (1, 2)$ .*
- (b) *If  $\nu \in (0, 1)$  then*

$$\mathfrak{G}_*^\varepsilon \in \mathcal{O}(\varepsilon^\nu), \quad \beta_*^\varepsilon - \mathfrak{J} \geq \mathcal{O}(\varepsilon^\nu), \quad \text{and} \quad \eta_*^\varepsilon(D^c) \in \mathcal{O}(\varepsilon^{2\nu \wedge (2-\nu)}). \quad (4.24)$$

*Proof.* Since  $2 - \nu < 2(\nu \wedge 1)$  for  $\nu \in [1, 2]$ , then, in view of Proposition 4.5, it suffices to prove that  $\eta_*^\varepsilon(\mathcal{N}) \in \mathcal{O}(\varepsilon^{2-\nu})$  for a bounded open neighborhood  $\mathcal{N}$  of  $z \in \mathcal{S} \setminus \mathcal{S}_s$ . Let  $\varphi$  be as in the proof of Lemma 3.6. By Proposition 4.5, we have  $\xi_2^\varepsilon \in \mathcal{O}(\varepsilon^{2(\nu \wedge 1)})$ , and  $\int_{B_r^c(\mathcal{S})} \varphi'(\mathcal{V}) \Delta \mathcal{V} d\eta_*^\varepsilon \in \mathcal{O}(\varepsilon^{2(\nu \wedge 1)})$ . Thus

$$\xi^\varepsilon \leq \frac{1}{2} \Delta \mathcal{V}(z) \eta_*^\varepsilon(B_r(z)) + \mathcal{O}(\varepsilon^{2(\nu \wedge 1)}) \quad (4.25)$$

by (3.35) and (3.40). Also  $\mathfrak{G}_*^\varepsilon \in \mathcal{O}(\varepsilon^{\nu \wedge 2})$  by Corollary 4.2 (a), which we combine with (4.25) and  $\nu \in (1, 2)$  to obtain

$$-C_0 \Delta \mathcal{V}(z) \eta_*^\varepsilon(B_r(z)) + \mathcal{O}(\varepsilon^2) \leq \varepsilon^{2-2\nu} \mathfrak{G}_*^\varepsilon \in \mathcal{O}(\varepsilon^{2-\nu})$$

by (3.42). Thus  $\eta_*^\varepsilon(B_r(z)) \in \mathcal{O}(\varepsilon^{2-\nu})$  for  $\nu \in (1, 2)$ . This completes the proof of part (a).

The proof of part (b) is divided in two steps.

Step 1. Suppose  $\mathfrak{J} = \mathfrak{J}_s$ . Then (4.8)–(4.11) hold with  $\mathfrak{J}$  replaced by  $\mathfrak{J}_s$ . By Lemma 3.5 we have  $\beta_*^\varepsilon - \mathfrak{J}_s \leq \mathcal{O}(\varepsilon^{\nu \vee (4\nu-2)})$ . Therefore  $\mathfrak{G}_*^\varepsilon \in \mathcal{O}(\varepsilon^\nu)$  by (4.11), and thus  $\tilde{\zeta}^\varepsilon \in \mathcal{O}(\varepsilon^\nu)$  by (4.10). Hence,  $\beta_*^\varepsilon - \mathfrak{J}_s \geq \mathcal{O}(\varepsilon^\nu)$  by (4.8). The estimate  $\eta_*^\varepsilon(D^c) \in \mathcal{O}(\varepsilon^{2\nu \wedge (2-\nu)})$  is obtained exactly as in Corollary 4.6 (a). Step 2. Suppose  $\mathfrak{J} < \mathfrak{J}_s$ . By Theorem 2.2 (ii), we may construct  $\mathcal{V}$  such that  $\mathcal{V}(z) > 5 \max_{\mathcal{S}_s} \mathcal{V}$  for all  $z \in \mathcal{S} \setminus \mathcal{S}_s$ . Let  $G = \{x \in \mathbb{R}^d : \mathcal{V}(x) < 2 \max_{\mathcal{S}_s} \mathcal{V}\}$  and  $\tilde{\varphi}$  be as in the proof of Lemma 4.1, with  $\delta = 2 \max_{\mathcal{S}_s} \mathcal{V}$ . We have

$$\mathfrak{J}_s - \ell(x) \leq \ell(z) - \ell(x) \leq C_\ell |x - z| \quad \forall z \in \mathcal{S}_s, \quad \text{and} \quad x \in \mathbb{R}^d.$$

Thus

$$\ell(x) - \mathfrak{J}_s \geq \max_{z \in \mathcal{S}_s} \{-C_\ell |x - z|\} = -C_\ell \text{dist}(x, \mathcal{S}_s) \quad \forall x \in \mathbb{R}^d.$$

Also by Proposition 4.5, for some positive constants  $r$  and  $\kappa_1$  we obtain

$$\int_{G^c} (\ell(x) - \mathfrak{J}_s) d\eta_*^\varepsilon \geq -\kappa_1 \sum_{z \in \mathcal{S} \setminus \mathcal{S}_s} \eta_*^\varepsilon(B_r(z)) + \mathcal{O}(\varepsilon^{2\nu}).$$

Therefore, splitting the integral over  $G$  and  $G^c$ , we obtain as in (4.7) that

$$\int_{\mathbb{R}^d} \ell d\eta_*^\varepsilon - \mathfrak{J}_s \geq -\kappa_1 \sum_{z \in \mathcal{S} \setminus \mathcal{S}_s} \eta_*^\varepsilon(B_r(z)) + \mathcal{O}(\varepsilon^{2\nu}) - \frac{C_\ell}{\sqrt{C_0}} \tilde{\zeta}^\varepsilon,$$

and since  $\tilde{\zeta}^\varepsilon \in \mathcal{O}(\varepsilon^\nu)$ , following the steps in (4.8)–(4.11) we have

$$-\kappa_1 \sum_{z \in \mathcal{S} \setminus \mathcal{S}_s} \eta_*^\varepsilon(B_r(z)) - \mathcal{O}(\varepsilon^{2\nu}) - \frac{C_\ell}{\sqrt{C_0}} \tilde{\zeta}^\varepsilon \leq \beta_*^\varepsilon - \mathfrak{J}_s, \quad (4.26)$$

and

$$\frac{1}{2} \mathcal{G}_*^\varepsilon \leq \beta_*^\varepsilon - \mathfrak{J}_s + \frac{C_\ell}{C_0^3} \varepsilon^2 + \frac{\sqrt{2} C_\ell}{C_0} \varepsilon^\nu \sqrt{|\tilde{\xi}^\varepsilon|} + \kappa_1 \sum_{z \in \mathcal{S} \setminus \mathcal{S}_s} \eta_*^\varepsilon(B_r(z)) + \mathcal{O}(\varepsilon^{2\nu}). \quad (4.27)$$

In view of (4.13) and (4.25) we have

$$\sum_{z \in \mathcal{S} \setminus \mathcal{S}_s} \eta_*^\varepsilon(B_r(z)) \leq \kappa_2 (\varepsilon^{2-2\nu} \mathcal{G}_*^\varepsilon + \varepsilon^{2\nu}) \quad (4.28)$$

for some positive constant  $\kappa_2$ . Since  $\beta_*^\varepsilon - \mathfrak{J}_s \leq \mathcal{O}(\varepsilon^{\nu \vee (4\nu-2)})$  by Lemma 3.6, and  $\nu < 1$ , combining (4.27) and (4.28) we obtain  $\mathcal{G}_*^\varepsilon \in \mathcal{O}(\varepsilon^\nu)$ . Therefore by (4.28), we obtain  $\eta_*^\varepsilon(B_r(z)) \in \mathcal{O}(\varepsilon^{2\nu \wedge (2-\nu)})$  for all  $z \in \mathcal{S} \setminus \mathcal{S}_s$ . In turn,  $\beta_*^\varepsilon - \mathfrak{J} \geq \mathcal{O}(\varepsilon^\nu)$  by (4.26). This completes the proof.  $\square$

*Remark 4.7.* If  $\nu = 1$  and  $\mathfrak{J}_c = \mathfrak{J}_s$ , then following the argument in Step 2 of the proof of Corollary 4.6 we obtain the same estimates as in (4.24). In this case we don't estimate  $\mathcal{G}_*^\varepsilon$  from (4.27), but rather use Corollary 4.2(a) which asserts that  $\mathcal{G}_*^\varepsilon \in \mathcal{O}(\varepsilon)$ . Thus  $\eta_*^\varepsilon(B_r(z)) \in \mathcal{O}(\varepsilon)$  by (4.28), which, in turn, implies that  $\beta_*^\varepsilon - \mathfrak{J} \geq \mathcal{O}(\varepsilon)$  by (4.26).

## 5. CONVERGENCE OF THE SCALED OPTIMAL STATIONARY DISTRIBUTIONS

We need the following definition.

**Definition 5.1.** For the rest of the paper  $\{\mathcal{B}_z : z \in \mathcal{S}\}$  is some collection of nonempty, disjoint balls, with each  $\mathcal{B}_z$  centered around  $z$ , and we define  $\mathcal{B}_{\mathcal{S}} := \cup_{z \in \mathcal{S}} \mathcal{B}_z$ .

Recall  $\widehat{V}_z^\varepsilon$  from Definition 4.3. For  $z \in \mathcal{S}$ , we define the ‘scaled’ density  $\widehat{\varrho}_z^\varepsilon(x) := \varepsilon^{\nu d} \varrho_*^\varepsilon(\varepsilon^\nu x + z)$ , and denote by  $\widehat{\eta}_z^\varepsilon$  the corresponding probability measure in  $\mathbb{R}^d$ . We also define the ‘normalized’ probability density  $\mathring{\varrho}_z^\varepsilon$  supported on  $\eta_*^\varepsilon(\mathcal{B}_z)$  by

$$\mathring{\varrho}_z^\varepsilon(x) := \begin{cases} \frac{\widehat{\varrho}_z^\varepsilon(x)}{\eta_*^\varepsilon(\mathcal{B}_z)} & \text{if } \varepsilon^\nu x + z \in \mathcal{B}_z, \\ 0 & \text{otherwise,} \end{cases}$$

and let  $\mathring{\eta}_z^\varepsilon(dx) = \mathring{\varrho}_z^\varepsilon(x) dx$ .

Section 5.1 which follows concerns the critical regime. The subcritical and supercritical regimes are treated in Section 5.2.

**5.1. Convergence to a Gaussian in the critical regime.** Recall the notation in Definitions 1.9 and 1.10. Also the scaled quantities in Definition 4.3. We start with the following lemma.

**Lemma 5.2.** *Assume  $\nu = 1$ . Fix any  $z \in \mathcal{S}$ . Then every sequence  $\varepsilon_n \searrow 0$  has a subsequence along which  $\widehat{V}_z^\varepsilon(x) - \widehat{V}_z^\varepsilon(z)$  converges to some  $\bar{V}_z \in \mathcal{C}^2(\mathbb{R}^d)$  uniformly on compact subsets of  $\mathbb{R}^d$ , and  $\beta_*^\varepsilon$  converges to some constant  $\bar{\beta}$ , and these satisfy*

$$\frac{1}{2} \Delta \bar{V}_z(x) + \langle M_z x, \nabla \bar{V}_z(x) \rangle - \frac{1}{2} |\nabla \bar{V}_z(x)|^2 = \bar{\beta} - \ell(z). \quad (5.1)$$

Moreover, for some constant  $\hat{c}_0$  we have

$$|\nabla \bar{V}_z(x)| \leq \hat{c}_0 (1 + |x|) \quad \forall \varepsilon \in (0, 1), \quad \forall x \in \mathbb{R}^d, \quad (5.2)$$

and

$$\bar{\beta} \leq A^+(M_z) + \ell(z). \quad (5.3)$$

*Proof.* If  $\nu = 1$ , then by (4.15) we obtain

$$\frac{1}{2}\Delta\widehat{V}_z^\varepsilon + \langle \widehat{m}_z^\varepsilon, \nabla\widehat{V}_z^\varepsilon \rangle - \frac{1}{2}|\nabla\widehat{V}_z^\varepsilon|^2 + \widehat{\ell}_z^\varepsilon = \beta_*^\varepsilon. \quad (5.4)$$

By applying [29, Lemma 5.1] to (5.4) and using the assumptions on the growth of  $m$  and  $\ell$ , it follows that there exists a constant  $\hat{c}_0$  such that

$$|\nabla\widehat{V}_z^\varepsilon(x)| \leq \hat{c}_0(1+|x|) \quad \forall \varepsilon \in (0,1), \quad \forall x \in \mathbb{R}^d. \quad (5.5)$$

It follows by (5.4) and the bound in (5.5) that  $\widehat{V}_z^\varepsilon$  is locally bounded in  $\mathcal{C}^{2,\alpha}(\mathbb{R}^d)$ , for any  $\alpha \in (0,1)$ . It is also clear that  $\widehat{m}_z^\varepsilon(x) \rightarrow M_z x$  and  $\widehat{\ell}_z^\varepsilon(x) \rightarrow \ell(z)$ , as  $\varepsilon \searrow 0$ , uniformly over compact sets. Thus, taking limits in (5.4) along some sequence  $\varepsilon_n \searrow 0$  we obtain a function  $\bar{V}_z \in \mathcal{C}^2(\mathbb{R}^d)$  and a constant  $\beta$  which satisfy (5.1). The bound in (5.2) follows by (5.5), while the bound in (5.3) follows by Theorem 1.18 (c).  $\square$

We fix some notation. The function  $\bar{V}_z$  for  $z \in \mathcal{S}$  denotes the limit obtained in Lemma 5.2. The associated ‘diffusion limit’, takes the form

$$d\bar{X}_t = (M_z \bar{X}_t - \nabla\bar{V}_z(\bar{X}_t)) dt + d\bar{W}_t, \quad (5.6)$$

and its extended generator is denoted by

$$\bar{\mathcal{L}}_z := \frac{1}{2}\Delta + \langle M_z x - \nabla\bar{V}_z(x), \nabla \rangle. \quad (5.7)$$

Since (5.3) holds for all  $z \in \mathcal{S}$ , then we must have  $\bar{\beta} \leq \mathfrak{J}_c$ , and Lemma 5.2 provides an alternate proof of the upper bound  $\limsup_{\varepsilon \searrow 0} \beta_*^\varepsilon \leq \mathfrak{J}_c$ , which was already shown in Lemma 3.3. In the next theorem we show that if  $\liminf_{\varepsilon_n \searrow 0} \eta_*^{\varepsilon_n}(\mathcal{B}_z) > 0$ , over some sequence  $\{\varepsilon_n\}$ , then the diffusion in (5.6) is positive recurrent.

**Theorem 5.3.** *Assume  $\nu = 1$ , and let  $\{\mathcal{B}_z : z \in \mathcal{S}\}$  be as in Definition 5.1. Let  $\varepsilon_n \searrow 0$  be any sequence satisfying  $\liminf_{n \rightarrow \infty} \eta_*^{\varepsilon_n}(\mathcal{B}_z) = \theta_z > 0$  for some  $z \in \mathcal{S}$ , and  $(\bar{V}_z, \bar{\beta}) \in \mathcal{C}^2(\mathbb{R}^d) \times \mathbb{R}$  be any limit point of  $(\widehat{V}_z^\varepsilon(x) - \widehat{V}_z^\varepsilon(z), \beta_*^\varepsilon)$  along some subsequence of  $\{\varepsilon_n\}$  (see Lemma 5.2). Recall Definition 1.9. Then*

- (a) *The diffusion in (5.6) is positive recurrent with invariant probability measure  $\bar{\eta}_z$ , and the density  $\hat{\varrho}_z^\varepsilon$  in Definition 5.1 converges to the density  $\bar{\varrho}_z$  of  $\bar{\eta}_z$ , uniformly on compact subsets of  $\mathbb{R}^d$ .*
- (b) *The invariant probability measure  $\bar{\eta}_z$  has finite second moments.*
- (c) *It holds that  $\bar{\beta} = \ell(z) + A^+(M_z)$ .*
- (d) *We have*

$$\widehat{V}_z(x) = \frac{1}{2}\langle x, \widehat{Q}_z x \rangle, \quad (5.8)$$

*and that  $\bar{\varrho}_z$  is the density of a Gaussian with mean 0 and covariance matrix  $\widehat{\Sigma}_z$ . Here  $(\widehat{Q}_z, \widehat{\Sigma}_z)$  are the pair of matrices which solve (1.16).*

- (e) *It holds that*

$$\liminf_{\varepsilon_n \searrow 0} \int_{\mathcal{B}_z} \left( \ell(x) + \frac{1}{2}|v_*^{\varepsilon_n}(x)|^2 \right) \eta_*^{\varepsilon_n}(dx) \geq \theta_z (\ell(z) + A^+(M_z)).$$

*Proof.* In order to show that the diffusion in (5.6) is positive recurrent, we examine the scaled diffusion

$$dX_t = (\widehat{m}_z^\varepsilon(X_t) - \nabla\widehat{V}_z^\varepsilon(X_t)) dt + dW_t. \quad (5.9)$$

Recall from Definition 5.1 that  $\hat{\eta}_z^\varepsilon$  and  $\hat{\varrho}_z^\varepsilon$  denote the invariant probability measure of (5.9) and its density, respectively. Let

$$\widehat{\mathcal{L}}_z^\varepsilon := \frac{1}{2}\Delta + \langle \widehat{m}_z^\varepsilon - \nabla\widehat{V}_z^\varepsilon, \nabla \rangle$$

denote the extended generator of (5.9). It follows by Lemma 4.1 and the Markov inequality that  $\eta_*^\varepsilon(\mathcal{B}_z \setminus B_{n\varepsilon}(z)) \leq \frac{\tilde{\kappa}_0}{n^2}$  for all  $n \in \mathbb{N}$ . Hence,  $\{\hat{\eta}_z^{\varepsilon_n} : n \in \mathbb{N}\}$  is a tight family of measures. By the Harnack inequality the family  $\{\hat{\varrho}_z^{\varepsilon_n} : n \in \mathbb{N}\}$  is locally bounded, and locally Hölder equicontinuous, and the same of course applies to  $\{\hat{\varrho}_z^{\varepsilon_n} : n \in \mathbb{N}\}$ . Moreover, the tightness of  $\{\hat{\eta}_z^{\varepsilon_n} : n \in \mathbb{N}\}$  implies the uniform integrability of  $\{\hat{\varrho}_z^{\varepsilon_n} : n \in \mathbb{N}\}$ . Select any subsequence, also denoted by  $\{\varepsilon_n\}$  along which  $\hat{\varrho}_z^{\varepsilon_n}$  converges locally uniformly, and denote the limit by  $\bar{\varrho}_z$ . By uniform integrability,  $\hat{\varrho}_z^{\varepsilon_n}$  also converges in  $L^1(\mathbb{R}^d)$ , as  $n \rightarrow \infty$ , and hence  $\int_{\mathbb{R}^d} \bar{\varrho}_z(x) dx = 1$ . Therefore  $\bar{\eta}_z(dx) := \bar{\varrho}_z(x) dx$  is a probability measure. Let  $f$  be a smooth function with compact support, and  $\bar{\mathcal{L}}_z$  be as in (5.7). Then

$$\begin{aligned} \left| \int_{\mathbb{R}^d} \hat{\mathcal{L}}_z^{\varepsilon_n} f(x) \hat{\varrho}_z^{\varepsilon_n}(x) dx - \int_{\mathbb{R}^d} \bar{\mathcal{L}}_z f(x) \bar{\varrho}_z(x) dx \right| &\leq \left| \int_{\mathbb{R}^d} \hat{\mathcal{L}}_z^{\varepsilon_n} f(x) (\hat{\varrho}_z^{\varepsilon_n}(x) - \bar{\varrho}_z(x)) dx \right| \\ &\quad + \left| \int_{\mathbb{R}^d} (\hat{\mathcal{L}}_z^{\varepsilon_n} f(x) - \bar{\mathcal{L}}_z f(x)) \bar{\varrho}_z(x) dx \right|. \end{aligned} \quad (5.10)$$

Since  $\hat{\varrho}_z^{\varepsilon_n} \rightarrow \bar{\varrho}_z$  in  $L^1(\mathbb{R}^d)$ , the first term on the right hand side of (5.10) converges to 0 as  $n \rightarrow \infty$ . Similarly, since  $\hat{m}_z^{\varepsilon_n}(x) \rightarrow M_z x$  and  $\nabla \hat{V}_z^{\varepsilon_n} \rightarrow \nabla \bar{V}_z$  uniformly on compact subsets of  $\mathbb{R}^d$ , the second term also converges to 0. Since  $\hat{\eta}_z^\varepsilon$  is an invariant probability measure of (5.9), by the definition of  $\hat{\varrho}_z^{\varepsilon_n}$  we have  $\int_{\mathbb{R}^d} \hat{\mathcal{L}}_z^{\varepsilon_n} f(x) \hat{\varrho}_z^{\varepsilon_n}(x) dx = 0$ , for all large enough  $n$ , which implies that  $\int_{\mathbb{R}^d} \bar{\mathcal{L}}_z f(x) \bar{\varrho}_z(x) dx = 0$ . Hence,  $\bar{\eta}_z$  is an infinitesimal invariant probability measure of (5.6), and since the diffusion is regular, it is also an invariant probability measure. This proves part (a).

Since the diffusion in (5.6) has an invariant probability measure, it follows that it is positive recurrent. By Lemma 4.1 we have

$$\sup_{\varepsilon \in (0,1)} \int_{\{\varepsilon^\nu x + z \in \mathcal{B}_z\}} |x|^2 \hat{\eta}_z^\varepsilon(dx) < \infty,$$

which implies by Fatou's lemma that  $\int_{\mathbb{R}^d} |x|^2 \bar{\eta}_z(dx) < \infty$ . Also by Theorem 1.4 and Theorem 1.18 (c) we must have  $\bar{\beta} - \ell(z) = \Lambda^+(M_z)$ . This completes the proof of parts (b) and (c).

By part (c) and Theorem 1.18 (c) the solution of (5.1) is unique and is given by (5.8). That  $\bar{\varrho}_z$  is Gaussian with covariance matrix  $\widehat{\Sigma}_z$  follows by the second equation in (1.16). This proves part (d).

Since  $\bar{V}_z$  has at most quadratic growth by (5.5), we have  $\int_{\mathbb{R}^d} |\bar{V}_z(x)| \bar{\eta}_z(dx) < \infty$ . Therefore, with  $\bar{\mathbb{E}}_x$  denoting the expectation operator for the process governed by (5.6), it is the case that  $\bar{\mathbb{E}}_x [\bar{V}_z(X_t)]$  converges as  $t \rightarrow \infty$  [21, Theorem 4.12]. Integrating both sides of (5.1) with respect to  $\bar{\eta}_z$ , we deduce that

$$\int_{\mathbb{R}^d} \frac{1}{2} |\nabla \bar{V}_z(x)|^2 \bar{\eta}_z(dx) = \bar{\beta} - \ell(z). \quad (5.11)$$

Using Fatou's lemma, we obtain by part (d) that

$$\begin{aligned} \liminf_{\varepsilon_n \searrow 0} \int_{\mathcal{B}_z} \mathcal{R}[v_*^{\varepsilon_n}](x) \eta_*^{\varepsilon_n}(dx) &= \liminf_{\varepsilon_n \searrow 0} \int_{\{\varepsilon^\nu x + z \in \mathcal{B}_z\}} \left( \hat{\ell}_z^{\varepsilon_n}(x) + \frac{1}{2} |\nabla \hat{V}_z^{\varepsilon_n}(x)|^2 \right) \hat{\eta}_z^{\varepsilon_n}(dx) \\ &\geq \lim_{R \rightarrow \infty} \liminf_{\varepsilon_n \searrow 0} \int_{\{|x| \leq R\}} \left( \hat{\ell}_z^{\varepsilon_n}(x) + \frac{1}{2} |\nabla \hat{V}_z^{\varepsilon_n}(x)|^2 \right) \eta_*^{\varepsilon_n}(\mathcal{B}_z) \hat{\eta}_z^{\varepsilon_n}(dx) \\ &\geq \theta_z (\Lambda^+(M_z) + \ell(z)), \end{aligned}$$

where in the second inequality we use (5.11), along with the hypothesis that  $\eta_*^{\varepsilon_n}(\mathcal{B}_z) \rightarrow \theta_z > 0$ . This proves part (e) and thus completes the proof.  $\square$

Part of the statement in Theorem 1.11 (iii) follows from the following result.

**Theorem 5.4.** *Recall the definition of  $\mathfrak{J}_c$  from Theorem 1.11. We assume  $\nu = 1$ . Then, it holds that  $\lim_{\varepsilon \searrow 0} \beta_*^\varepsilon = \mathfrak{J}_c$ . Also  $\bar{\beta}$  in (5.1) equals  $\mathfrak{J}_c$ . Moreover, for any  $r > 0$  we have*

$$\lim_{\varepsilon \searrow 0} \eta_*^\varepsilon(B_r^c(\mathcal{Z}_c)) = 0, \quad \text{and} \quad \lim_{\varepsilon \searrow 0} \int_{B_r^c(\mathcal{Z}_c)} |v_*^\varepsilon(x)|^2 \eta_*^\varepsilon(dx) = 0. \quad (5.12)$$

*Proof.* Since the collection  $\{\mathcal{B}_z\}$  used in Theorem 5.3 was arbitrary, without loss of generality, we may let  $\mathcal{B}_z = B_r(z)$ . Let  $\varepsilon_n \searrow 0$  be any sequence such that  $\eta_*^{\varepsilon_n}(B_r(z)) \rightarrow \theta_z$  for all  $z \in \mathcal{S}$ , and define  $\mathcal{S}_o := \{z \in \mathcal{S} : \theta_z > 0\}$ . Thus  $\sum_{z \in \mathcal{S}_o} \theta_z = 1$ . By Theorem 5.3 (e) we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \beta_*^{\varepsilon_n} &\geq \sum_{z \in \mathcal{S}_o} \int_{B_r(z)} \left( \ell(x) + \frac{1}{2} |v_*^{\varepsilon_n}(x)|^2 \right) \eta_*^{\varepsilon_n}(dx) \\ &\geq \sum_{z \in \mathcal{S}_o} \theta_z \left( \ell(z) + \Lambda^+(Dm(z)) \right) \geq \mathfrak{J}_c. \end{aligned} \quad (5.13)$$

Since  $\limsup_{\varepsilon \searrow 0} \beta_*^\varepsilon \leq \mathfrak{J}_c$  by Lemma 3.3, (5.13) implies that  $\lim_{\varepsilon \searrow 0} \beta_*^\varepsilon = \mathfrak{J}_c$ . By Lemma 5.2 we have  $\liminf_{\varepsilon \searrow 0} \beta_*^\varepsilon \leq \bar{\beta}$ , and  $\bar{\beta} \leq \mathfrak{J}_c$  by (5.3). Therefore  $\bar{\beta} = \mathfrak{J}_c$ .

Given any sequence  $\varepsilon_n \searrow 0$ , we can extract a subsequence also denoted by  $\{\varepsilon_n\}$  along which  $\lim_{n \rightarrow \infty} \eta_*^{\varepsilon_n}(B_r(z)) \rightarrow \theta_z$  for all  $z \in \mathcal{S}$ . Then (5.13) holds. Also, by Proposition 4.5 we have  $\int_{B_r^c(z)} \ell(x) \eta_*^\varepsilon(dx) \rightarrow 0$  as  $\varepsilon \searrow 0$ . It is then clear that both assertions in (5.12) follow by (5.13).  $\square$

It is interesting to note that (5.8) holds for any  $z \in \mathcal{Z}_c$  even if  $\lim_{\varepsilon \rightarrow 0} \eta_*^{\varepsilon_n}(\mathcal{B}_z) = 0$ . This is part of the corollary that follows.

**Corollary 5.5.** *Suppose  $\nu = 1$ . Then for any  $z \in \mathcal{Z}_c$ , we have*

$$\widehat{V}_z^\varepsilon(x) - \widehat{V}_z^\varepsilon(z) \xrightarrow[\varepsilon \searrow 0]{} \frac{1}{2} \langle x, \widehat{Q}_z x \rangle,$$

*uniformly on compact sets. Also, unless  $z \in \mathcal{Z}_c$ , then the family  $\{\widehat{\eta}_z^\varepsilon : \varepsilon \in (0, 1)\}$  is not tight.*

*Proof.* Since  $\bar{\beta}$  in (5.1) equals  $\mathfrak{J}_c$  by Theorem 5.4, then, provided  $z \in \mathcal{Z}_c$ , the right hand side of (5.1) equals  $\Lambda^+(M_z)$ . The first assertion then follows by Theorem 1.18 (c).

If the family  $\{\widehat{\eta}_z^\varepsilon : \varepsilon \in (0, 1)\}$  is tight, then it follows from the proof of Theorem 5.3 that the diffusion limit in (5.6) is positive recurrent. However, if  $z \notin \mathcal{Z}_c$ , then  $\bar{\beta} - \ell(z) = \mathfrak{J}_c - \ell(z) < \Lambda^+(M_z)$ , and by the results of Theorem 1.4 and Theorem 1.18 (c), the diffusion in (5.6) has to be transient. Therefore,  $\{\widehat{\eta}_z^\varepsilon\}$  cannot be tight.  $\square$

**Remark 5.6.** It is worth examining the diffusion in (5.6) in the context of Example 1.14. Consider the example with the first set of data, and let  $c = 5$ . Then  $\mathfrak{S} = \{0\}$  and  $\mathfrak{J}_c = 2$ . Thus, for  $z = 0$ , we have  $\bar{V}_z = \bar{V}_0 = 2x^2$ , and the drift in (5.6) equals  $-2\bar{X}_t$ . For  $z = -1$ , we have  $\ell(-1) = 5$ ,  $Dm(-1) = -3$ , and direct substitution shows that  $\bar{V}_{-1} = -3x^2$  solves (5.1). The associated diffusion in (5.6) has drift  $3\bar{X}_t$ , and thus it is transient.

**5.2. Convergence to a Gaussian in the subcritical and supercritical regimes.** We return to the analysis of the subcritical and supercritical regimes in order to determine the asymptotic behavior of the density of the optimal stationary distribution in the vicinity of the stochastically stable set. In these regimes there are two scales. If we center the coordinates around a point in  $\mathfrak{S}$ , then we have  $V^\varepsilon(x) \in \mathcal{O}(\varepsilon^{-2}|x|^2)$ , and  $-\log \varrho_*^\varepsilon(x) \in \mathcal{O}(\varepsilon^{-2\nu}|x|^2)$ . To avoid this incompatibility we use the function  $\tilde{V}_z(x) = \varepsilon^{2(1-\nu)} V^\varepsilon(\varepsilon^\nu x)$  in the analysis, which scales correctly in space for all  $\nu$ . We have the following result.

**Theorem 5.7.** *Assume  $\nu \in (0, 2)$  and let  $\{\mathcal{B}_z : z \in \mathcal{S}\}$  be as in Definition 5.1. The following hold.*

- (a) Suppose that for some  $z \in \mathcal{S}$  and a sequence  $\varepsilon_n \searrow 0$  it holds that  $\liminf_{n \rightarrow \infty} \eta_*^{\varepsilon_n}(\mathcal{B}_z) > 0$ . Then the density  $\hat{\varrho}_z^{\varepsilon_n}$  in Definition 5.1 converges as  $n \rightarrow \infty$  (uniformly on compact sets) to the density of a Gaussian with mean 0 and covariance matrix  $\widehat{\Sigma}_z$  given in (1.16).
- (b) If  $\nu \in (1, 2)$  and  $z \in \mathcal{S} \setminus \widetilde{\mathcal{Z}}$ , then  $\lim_{\varepsilon \searrow 0} \eta_*^{\varepsilon}(\mathcal{B}_z) = 0$ .

*Proof.* The proof closely follows those of Lemma 5.2 and Theorem 5.3. Only the scaling differs. We summarize the essential steps.

First, suppose  $\nu < 1$ . Since  $\liminf_{n \rightarrow \infty} \eta_*^{\varepsilon_n}(\mathcal{B}_z) > 0$  then necessarily  $z \in \mathcal{S}_s$  by Lemma 3.6. We scale the space as  $1/\varepsilon^\nu$ , and use (4.15) which we write again here:

$$\frac{1}{2}\Delta\check{V}_z^\varepsilon(x) + \langle \hat{m}_z^\varepsilon(x), \nabla\check{V}_z^\varepsilon(x) \rangle - \frac{1}{2}|\nabla\check{V}_z^\varepsilon(x)|^2 = \varepsilon^{2(1-\nu)}(\beta_*^\varepsilon - \hat{\ell}_z^\varepsilon(x)). \quad (5.14)$$

By Lemma 4.4,  $\nabla\check{V}_z^\varepsilon = \varepsilon^{2(1-\nu)}\nabla\widehat{V}_z^\varepsilon$  is locally bounded and has at most linear growth. We write (5.14) as a HJB equation

$$\frac{1}{2}\Delta\check{V}_z^\varepsilon(x) + \min_{\bar{u} \in \mathbb{R}^d} \left[ \langle \hat{m}_z^\varepsilon(x) + \bar{u}, \nabla\check{V}_z^\varepsilon(x) \rangle + \frac{1}{2}|\bar{u}|^2 \right] = \varepsilon^{2(1-\nu)}(\beta_*^\varepsilon - \hat{\ell}_z^\varepsilon(x)). \quad (5.15)$$

The associated scaled controlled diffusion is

$$d\hat{X}_t = (\hat{m}_z^\varepsilon(\hat{X}_t) - \check{U}_t) dt + d\hat{W}_t. \quad (5.16)$$

Taking limits in (5.15) along some subsequence  $\varepsilon_n \searrow 0$ , we obtain a function  $\bar{V}_z \in \mathcal{C}^2(\mathbb{R}^d)$  of at most quadratic growth satisfying

$$\frac{1}{2}\Delta\bar{V}_z(x) + \min_{\bar{u} \in \mathbb{R}^d} \left[ \langle M_z x + \bar{u}, \nabla\bar{V}_z(x) \rangle + \frac{1}{2}|\bar{u}|^2 \right] = 0. \quad (5.17)$$

The associated diffusion limit is

$$d\bar{X}_t = (M_z \bar{X}_t - \nabla\bar{V}_z(\bar{X}_t)) dt + d\bar{W}_t. \quad (5.18)$$

As in Section 5.1,  $\hat{\eta}_*^\varepsilon$  denotes the invariant probability measure of (5.16) under the control  $\check{U}_t = -\nabla\check{V}_z^\varepsilon(X_t)$ , and  $\hat{\varrho}_*^\varepsilon$  its density. Following the proof of Theorem 5.3, and using Lemma 4.1, we deduce that the density  $\hat{\varrho}_z^\varepsilon$  in Definition 5.1 converges as  $\varepsilon_n \searrow 0$  to the density  $\bar{\varrho}_z$  of the invariant probability measure of (5.18). However since  $M_z$  is Hurwitz, then  $\Lambda^+(M_z) = 0$ , and by Theorem 1.18 we obtain  $\bar{V}_z = 0$ . So in this case (5.17) is trivial, and the covariance matrix  $\widehat{\Sigma}_z$  of the Gaussian is the solution of (1.16) with  $\widehat{Q}_z = 0$ .

Next we assume  $\nu \in (1, 2)$ , and we use the same scaling and definitions as for the subcritical regime, except that  $z \in \mathcal{Z}$ . It is clear that

$$\varepsilon^{2(1-\nu)}(\hat{\ell}_z^\varepsilon(x) - \ell(z)) \leq C_\ell \varepsilon^{2(1-\nu)} \varepsilon^\nu |x| \xrightarrow[\varepsilon \searrow 0]{} 0,$$

where  $C_\ell$  denotes a Lipschitz constant of  $\ell$ . By Corollary 4.2 the constants  $\varepsilon^{2(1-\nu)}(\beta_*^\varepsilon - \ell(z))$  are bounded, uniformly in  $\varepsilon \in (0, 1)$ . Therefore, as argued in the proof of Theorem 5.3, for every sequence  $\varepsilon_n \searrow 0$ , there exists a subsequence, also denoted as  $\{\varepsilon_n\}$  along which  $\varepsilon_n^{2(1-\nu)}(\beta_*^\varepsilon - \ell(x))$  converges to a constant  $\hat{\beta}$ , and  $\check{V}_z^\varepsilon(\cdot) - \check{V}_z^\varepsilon(z)$  converges to some  $\bar{V}_z \in \mathcal{C}^2(\mathbb{R}^d)$ , uniformly on compact sets. Taking limits in (5.15) along this subsequence, we obtain

$$\frac{1}{2}\Delta\bar{V}_z(x) + \min_{\bar{u} \in \mathbb{R}^d} \left[ \langle M_z x + \bar{u}, \nabla\bar{V}_z(x) \rangle + \frac{1}{2}|\bar{u}|^2 \right] = \hat{\beta}. \quad (5.19)$$

Recall the notation  $\widetilde{\mathcal{Z}}$  and  $\widetilde{\mathfrak{J}}$  in Definition 1.10. By Lemma 3.3 we have

$$\hat{\beta} \leq \widetilde{\mathfrak{J}} = \min_{z \in \mathcal{Z}} \Lambda^+(Dm(z)). \quad (5.20)$$

Following exactly the same steps as in the proof of Theorem 5.3, we deduce that the diffusion in (5.18) is positive recurrent, with an invariant probability measure  $\bar{\eta}_z$  that has finite second

moments, and that the density  $\varrho_z^\varepsilon$  in Definition 5.1 converges as  $\varepsilon_n \searrow 0$  to the density  $\bar{\varrho}_z$  of  $\bar{\eta}_z$ . Therefore,

$$\Lambda^+(Dm(z)) = \hat{\beta} \quad (5.21)$$

by Theorem 1.18(c). Thus  $\hat{\beta} = \tilde{\mathfrak{J}} = \Lambda^+(Dm(z))$  by (5.20)–(5.21). This shows that unless  $z \in \tilde{\mathcal{Z}}$ , the hypothesis  $\liminf_{n \rightarrow \infty} \eta_*^{\varepsilon_n}(\mathcal{B}_z) > 0$  cannot hold, thus establishing part (b) of the theorem.

With  $z \in \tilde{\mathcal{Z}}$ , and  $\hat{\beta} = \tilde{\mathfrak{J}}$ , equation (5.19) has a unique solution by Theorem 1.18(c), and we obtain  $\bar{V}_z(x) = \frac{1}{2}\langle x, \hat{Q}_z \rangle$ , and that  $\bar{\varrho}_z$  is the density of a Gaussian with mean 0 and covariance matrix  $\hat{\Sigma}_z$ , with  $(\hat{Q}_z, \hat{\Sigma}_z)$  the pair of matrices which solve (1.16). This completes the proof.  $\square$

## 6. CONCLUDING REMARKS

In general, Morse–Smale flows may contain hyperbolic closed orbits, and it would be desirable to extend the results of the paper accordingly. An energy function  $\mathcal{V}$  as in Theorem 2.2 may be constructed to account for critical elements that are closed orbits [30, 38]. Note that under the control used in Remark 3.4 the optimal stationary distribution concentrates on the minimum of  $\mathcal{V}$ . In the case that  $z \in \mathbb{R}^d$  belongs to a stable periodic orbit with period  $T_0$ , we can construct  $\mathcal{V}$  so that it attains its minimum on this closed orbit. In this manner, if  $\phi_t$  denotes the flow of the vector field  $m$ , then it follows by (3.8) that under the control used in Remark 3.4, we obtain

$$\int_{\mathbb{R}^d} \ell(x) \mu^\varepsilon(dx) \xrightarrow[\varepsilon \searrow 0]{} \frac{1}{T_0} \int_0^{T_0} \ell(\phi_t(z)) dt.$$

The same can be done in the subcritical regime, by modifying the proof of Lemma 3.5, and using instead the approach in Remark 3.4. We leave it up to the reader to verify that Lemma 3.1 still holds if the set of critical elements  $\mathcal{S}$  contains hyperbolic closed orbits. Let us define

$$\dot{\ell}(z) := \frac{1}{T_0} \int_0^{T_0} \ell(\phi_t(z)) dt,$$

when  $z$  belongs to a closed orbit, and  $\dot{\ell}(z) = \ell(z)$ , when  $m(z) = 0$ . Then, provided  $\text{Arg min}_{z \in \mathcal{S}} \dot{\ell}(z)$  contains only stable critical elements, then the support of the limit of the optimal stationary distribution lies in  $\mathcal{S}_s$ , and this is true in any of the three regimes. However, the full analysis when unstable closed orbits are involved seems to be more difficult.

## APPENDIX A. PROOFS OF THE RESULTS IN SECTION 1.2

We start with the proof of Lemma 1.3.

*Proof of Lemma 1.3.* The proof is standard. Let  $U$  be given and define  $M_t := \mathbb{E}[ \int_0^t |U_s|^2 ds ]$ ,  $t \in \mathbb{R}_+$ . For  $T > 0$ , let  $\mathcal{H}_T^2$  denote the space of  $\{\mathfrak{F}_t\}$ -adapted processes  $Y$  defined on  $[0, T]$ , having continuous sample paths, and satisfying  $\mathbb{E}[\sup_{0 \leq t \leq T} |Y_t|^2] < \infty$ . The space  $\mathcal{H}_T^2$  (more precisely the set of equivalence classes in  $\mathcal{H}_T^2$ ) is a Banach space under the norm

$$\|Y\|_{\mathcal{H}_T^2} := \left( \mathbb{E} \left[ \sup_{0 \leq t \leq T} |Y_t|^2 \right] \right)^{1/2}.$$

It is standard to show, for example following the proof of [1, Theorem 2.2.2] that any solution  $X$  of (1.1) satisfies

$$\|X - X_0\|_{\mathcal{H}_t^2}^2 \leq \kappa_0 t (1+t) (1 + M_t + \mathbb{E}[|X_0|^2]) e^{\kappa_1 t} \quad \forall t \geq 0,$$

for some constants  $\kappa_0$  and  $\kappa_1$  that depend only on  $m$ . The existence of a pathwise unique solution then follows by applying the contraction mapping theorem as in [1, Theorem 2.2.4].  $\square$

The rest of this section is devoted to the proof of Theorem 1.4. Without loss of generality we fix  $\varepsilon = 1$ , and denote the optimal value  $\beta_*^\varepsilon$  in (1.4) as  $\beta_*$ . Also throughout the rest of this section, without loss of generality we assume that  $\ell \geq 0$ .

We proceed by establishing two key lemmas, followed by the proof of Theorem 1.4. Recall the definition of  $\mathcal{R}$  in (1.2). For  $x \in \mathbb{R}^d$ , and  $\alpha > 0$ , we define the subset  $\mathfrak{U}_x^\alpha$  of admissible controls by

$$\mathfrak{U}_x^\alpha := \left\{ U \in \mathfrak{U} : \mathbb{E}_x^U \left[ \int_0^\infty e^{-\alpha s} \mathcal{R}(X_s, U_s) ds \right] < \infty \right\}, \quad (\text{A.1})$$

where  $\mathbb{E}_x^U$  denotes the expectation under the law of  $(X, U)$ , with  $X_0 = x$  for the solution of

$$X_t = x + \int_0^t m(X_s) ds + \int_0^t U_s ds + W_t, \quad t \geq 0. \quad (\text{A.2})$$

**Lemma A.1.** *The equation*

$$\frac{1}{2} \Delta V_\alpha + \langle m, \nabla V_\alpha \rangle - \frac{1}{2} |\nabla V_\alpha|^2 + \ell = \alpha V_\alpha \quad (\text{A.3})$$

has a solution in  $\mathcal{C}^2(\mathbb{R}^d)$  for all  $\alpha \in (0, 1)$ . Moreover, for all  $\alpha \in (0, 1)$ , we have the following:

(i) For some constant  $c_0 > 0$ , not depending on  $\alpha$ , it holds that

$$|\nabla V_\alpha(x)| \leq c_0 \sqrt{1 + |x|}, \quad \text{and} \quad |\alpha V_\alpha(x)| \leq \ell(x) + \frac{c_0}{\alpha} \quad \forall x \in \mathbb{R}^d. \quad (\text{A.4})$$

(ii) The function  $V_\alpha$  satisfies

$$V_\alpha(x) \leq \inf_{U \in \mathfrak{U}_x^\alpha} \mathbb{E}_x^U \left[ \int_0^\infty e^{-\alpha s} \mathcal{R}(X_s, U_s) ds \right], \quad \forall x \in \mathbb{R}^d. \quad (\text{A.5})$$

(iii) With  $\bar{c}_\ell$  the constant in (1.8), we have

$$\inf_{\{x : \ell(x) \leq \bar{c}_\ell\}} \alpha V_\alpha = \inf_{\mathbb{R}^d} \alpha V_\alpha \leq \bar{c}_\ell.$$

*Proof.* In [2, Theorem 4.18, p. 177] it is proved that (A.3) has a solution in  $\mathcal{C}^2(\mathbb{R}^d)$ , and it also shown in the proof of this theorem that there exists a constant  $\kappa_0 > 0$  which does not depend on  $\alpha$  such that

$$\alpha V_\alpha(x) \geq -\kappa_0 \quad \forall x \in \mathbb{R}^d. \quad (\text{A.6})$$

By [21, Theorem B.1] there exists a constant  $C$  not depending on  $R > 0$  such that

$$\sup_{B_R} |\nabla V_\alpha| \leq C \left( 1 + \sup_{B_{R+1}} \sqrt{(\alpha V_\alpha)^-} + \sup_{B_{R+1}} \sqrt{\ell^+} + \sup_{B_{R+1}} |\nabla \ell|^{1/3} \right).$$

from which gradient estimate in (A.4) follows. The structural assumption on the Hamiltonian  $h(x, p)$  in [21, Theorem B.1] is  $p \mapsto h(x, p)$  is strictly convex for all  $x \in \mathbb{R}^d$ , and there exists some constant  $k_0 > 0$  such that

$$k_0 |p|^2 \leq h(x, p) \leq k_0^{-1} |p|^2, \quad |\nabla_x h(x, p)| \leq k_0^{-1} (1 + |p|^2), \quad (x, p) \in \mathbb{R}^{2d}. \quad (\text{A.7})$$

This Hamiltonian corresponds to  $h(x, p) = \frac{1}{2} |p|^2 - \langle m, p \rangle$  for the equation in (A.3), and the first bound in (A.7) is not satisfied. However, replacing this bound with

$$k_0 (|p|^2 - k_1) \leq h(x, p) \leq k_0^{-1} (|p|^2 + k_1),$$

for some constant  $k_1 \geq 0$ , the proof of [21, Theorem B.1] goes through unmodified.

Recall the definition of  $\widehat{\mathfrak{U}}$  in (1.9). Writing (A.3) in HJB form, and applying Itô's formula we obtain

$$V_\alpha(x) - e^{-\alpha t} \mathbb{E}_x^U [V_\alpha(X_t)] \leq \mathbb{E}_x^U \left[ \int_0^t e^{-\alpha s} \mathcal{R}(X_s, U_s) ds \right] \quad \forall t > 0, \quad \forall U \in \widehat{\mathfrak{U}}. \quad (\text{A.8})$$

Since  $m$  is bounded, then it is standard to show using (A.2) that

$$\mathbb{E}_x^U \left[ \sup_{0 \leq s \leq t} |X(s) - x| \right] \leq \|m\|_\infty t + \sqrt{t} + \mathbb{E}_x^U \left[ \int_0^t |U_s| ds \right] < \infty \quad \forall U \in \widehat{\mathfrak{U}}, \quad \forall t > 0. \quad (\text{A.9})$$

Also, if  $\mathbb{E}_x^0$  denotes the expectation  $\mathbb{E}_x^U$  with  $U = 0$ , then by (A.2) we have the estimate

$$\mathbb{E}_x^0 [|X_t|^2] \leq \kappa_2 (1 + t^2 + |x|^2) < \infty \quad \forall t > 0, \quad (\text{A.10})$$

for some constant  $\kappa_2$ . As shown in the proof of [2, Theorem 4.18, p. 177],  $\alpha \mapsto \alpha V_\alpha(0)$  is bounded on  $(0, 1)$ , which together with the gradient estimate in (A.4) we have already proved, provides us with a liberal bound of  $V_\alpha$  of the form  $|V_\alpha(x)| \leq C(1 + |x|^2)$  for some constant  $C$ . This combined with (A.10) implies that  $e^{-\alpha t} \mathbb{E}_x^0 [V_\alpha(X_t)] \rightarrow 0$  as  $t \rightarrow \infty$ . Therefore, using (A.9), and the Lipschitz constant  $C_\ell$  of  $\ell$ , we obtain by (A.8) that

$$\begin{aligned} \alpha V_\alpha(x) &\leq \mathbb{E}_x^0 \left[ \int_0^\infty \alpha e^{-\alpha s} \ell(X_s) ds \right] \\ &\leq \ell(x) + C_\ell \int_0^\infty \alpha e^{-\alpha s} (\|m\|_\infty s + 2\sqrt{s}) ds \quad \forall x \in \mathbb{R}^d, \end{aligned}$$

which results in the estimate given in (A.4), where without loss of generality we use a common constant  $c_0$ . This completes the proof of part (i).

Let  $g(x, t) := |x| + \|m\|_\infty t + 2\sqrt{t}$ . Multiplying both sides of (A.9) by  $e^{-\alpha t}$ , strengthening the inequality, and applying the Hölder inequality, we obtain

$$\begin{aligned} e^{-\alpha t} \mathbb{E}_x^U [|X_t|] &\leq g(x, t) e^{-\alpha t} + e^{-\frac{\alpha}{2} t} \mathbb{E}_x^U \left[ \int_0^t e^{-\frac{\alpha}{2} s} |U_s| ds \right] \\ &\leq g(x, t) e^{-\alpha t} + \sqrt{t} e^{-\frac{\alpha}{4} t} \left( \mathbb{E}_x^U \left[ \int_0^t e^{-\alpha s} |U_s|^2 ds \right] \right)^{1/2} \xrightarrow[t \rightarrow \infty]{} 0 \end{aligned} \quad (\text{A.11})$$

for all  $U \in \mathfrak{U}_x^\alpha$ , as defined in (A.1). Taking limits as  $t \rightarrow \infty$  in (A.8), and using (A.11), and the bound of  $V_\alpha$  in (A.4) together with  $|\ell(x)| \leq C_l|x| + |\ell(0)|$ , we obtain (A.5).

We now turn to part (iii). Let

$$\chi(x) := \frac{1}{\sqrt{3}} \left( \min_{y \in B_1(x)} [\ell(y) - (d + 1 + 2\sqrt{d} \|m\|_\infty)^2] \right)^{1/2},$$

and

$$\psi(x) := V_\alpha(x) + \frac{2\kappa_0}{\alpha} - \chi(x_0)(1 - |x - x_0|^2), \quad x \in B_1(x_0),$$

where  $\kappa_0 > 0$  is the constant in (A.6). With  $\phi(x) := |x - x_0|^2$ , we have

$$\begin{aligned} -\frac{1}{2} \Delta \psi - \langle m - \nabla V_\alpha, \nabla \psi \rangle + \alpha \psi &= \left( -\frac{1}{2} \Delta V_\alpha - \langle m, \nabla V_\alpha \rangle + \frac{1}{2} |\nabla V_\alpha|^2 - \alpha V_\alpha \right) \\ &\quad + \frac{1}{2} |\nabla V_\alpha - \chi(x_0) \nabla \phi|^2 - 2\chi^2(x_0) \phi + 2\kappa_0 \\ &\quad - \chi(x_0) \left( \frac{1}{2} \Delta \phi + \langle m, \nabla \phi \rangle + \alpha(1 - \phi) \right) \\ &\geq \ell - 2\chi^2(x_0) + 2\kappa_0 - (d + 2\sqrt{d} \|m\|_\infty + 1) \chi(x_0) \\ &\geq \ell - 3\chi^2(x_0) - (d + 1 + 2\sqrt{d} \|m\|_\infty)^2 \quad \text{in } B_1(x_0), \quad \forall \alpha \in (0, 1), \end{aligned}$$

where we use (A.3) and the fact that  $\kappa_0 \geq 0$ . Since  $\psi > 0$  on  $\partial B_1(x_0)$  by (A.6), an application of the strong maximum principle shows that  $\psi \geq 0$  in  $B_1(x_0)$ , which implies that

$$\alpha V_\alpha(x) \geq \alpha \chi(x) + \kappa_0 \quad \forall x \in \mathbb{R}^d.$$

Since  $\ell$  is inf-compact, and therefore the same is true for  $\chi$  by its definition, this shows that  $\alpha V_\alpha$  is inf-compact. In particular, it attains its infimum in  $\mathbb{R}^d$ . With  $\eta_0$  denoting the invariant probability measure of the diffusion in (A.2) under the control  $U = 0$ , using (A.5), we obtain

$$\inf_{\mathbb{R}^d} V_\alpha \leq \int_{\mathbb{R}^d} V_\alpha \, d\eta_0 \leq \int_{\mathbb{R}^d} \mathbb{E}_x \left[ \int_0^\infty e^{-\alpha s} \ell(X_s) \, ds \right] \eta_0(dx) \leq \frac{\bar{c}_\ell}{\alpha}, \quad (\text{A.12})$$

where the last inequality follows by (1.8). One more application of the maximum principle implies that if  $V_\alpha$  attains its infimum at  $\hat{x} \in \mathbb{R}^d$  then  $\ell(\hat{x}) \leq \alpha V_\alpha(\hat{x})$ . This together with (A.12) implies part (iii).  $\square$

*Remark A.2.* We should mention, even though we don't need it for the proof of the main theorem, that (A.5) holds with equality, and thus  $V_\alpha$  is indeed the value of the infinite horizon discounted control problem. The proof of this assertion goes as follows. Since  $\nabla V_\alpha$  has at most linear growth, the diffusion in (A.2) under the Markov control  $v_\alpha = -\nabla V_\alpha$  has a unique strong solution. It is also clear by (A.4) that for any  $\alpha > 0$  we can select a constant  $\kappa_1(\alpha)$  such that  $|\nabla V_\alpha(x)| \leq \kappa_1(\alpha) + \frac{\alpha}{16}x$ . Thus using a standard estimate [1, Theorem 2.2.2] we obtain

$$\mathbb{E}_x^{v_\alpha} \left[ \sup_{0 \leq s \leq t} |X(s)|^2 \right] \leq \kappa_2(\alpha)(1+t^2)(1+|x|^2)e^{\frac{\alpha}{2}t} \quad (\text{A.13})$$

for some constant  $\kappa_2(\alpha) > 0$ . With  $\tau_R$  denoting the first exit time from  $B_R$ , applying Dynkin's formula we obtain

$$V_\alpha(x) = \mathbb{E}_x^{v_\alpha} \left[ \int_0^{t \wedge \tau_R} e^{-\alpha s} \mathcal{R}(X_s, v_\alpha(X_s)) \, ds \right] + \mathbb{E}_x^{v_\alpha} [e^{-\alpha(t \wedge \tau_R)} V_\alpha(X_{t \wedge \tau_R})].$$

We write

$$\mathbb{E}_x^{v_\alpha} [e^{-\alpha(t \wedge \tau_R)} V_\alpha(X_{t \wedge \tau_R})] = A_1(t, R) + A_2(t, R).$$

with

$$A_1(t, R) := \mathbb{E}_x^{v_\alpha} [e^{-\alpha t} V_\alpha(X_{t \wedge \tau_R}) \mathbf{1}_{\{t \leq \tau_R\}}], \quad A_2(t, R) := \mathbb{E}_x^{v_\alpha} [e^{-\alpha \tau_R} V_\alpha(X_{t \wedge \tau_R}) \mathbf{1}_{\{\tau_R < t\}}].$$

Since  $V_\alpha$  has at most linear growth in  $x$  by (A.4), it follows by (A.13) that

$$\lim_{t \rightarrow \infty} \limsup_{R \rightarrow \infty} |A_1(t, R)| = 0.$$

We also have  $\limsup_{R \rightarrow \infty} |A_2(t, R)| = 0$  by dominated convergence, since  $\mathbb{P}_x^{v_\alpha}(\tau_R < t) \rightarrow 0$  as  $R \rightarrow \infty$ . Thus, taking limits first as  $R \rightarrow \infty$ , and then as  $t \rightarrow \infty$  in (A.13), we obtain

$$V_\alpha(x) \geq \mathbb{E}_x^{v_\alpha} \left[ \int_0^\infty e^{-\alpha s} \mathcal{R}(X_s, v_\alpha(X_s)) \, ds \right].$$

Thus the converse inequality to (A.5) also holds.

Recall the definition of  $\beta_*^\varepsilon$  in (1.4), which for  $\varepsilon = 1$  we denote simply as  $\beta_*$ . Also define the class of controls  $\overline{\mathfrak{U}}_x$  by

$$\overline{\mathfrak{U}}_x := \left\{ U \in \mathfrak{U} : \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_x^U \left[ \int_0^T \mathcal{R}(X_s, U_s) \, ds \right] < \infty \right\}.$$

**Lemma A.3.** *There exists  $V \in \mathcal{C}^2(\mathbb{R}^d)$  which is bounded below in  $\mathbb{R}^d$  and satisfies*

$$\mathcal{A}[V](x) := \frac{1}{2} \Delta V + \langle m, \nabla V \rangle - \frac{1}{2} |\nabla V|^2 + \ell = \beta. \quad (\text{A.14})$$

for  $\beta = \beta_*$ . Moreover, under the Markov control  $U_t = -v_*(X_t)$ , with  $v_* = -\nabla V$ , the diffusion in (A.2) is positive recurrent, and  $\beta_* = \int_{\mathbb{R}^d} \mathcal{R}[v_*](x) \, d\eta_*$ , where  $\eta_*$  is the invariant probability measure corresponding to the control  $v_*$ .

*Proof.* The existence of this solution is established as a limit of  $V_\alpha(\cdot) - V_\alpha(0)$ , with  $V_\alpha$  the solution of (A.3) in Lemma A.1 along some sequence  $\alpha_n \searrow 0$  [2, p. 175]. It also follows from the proof from this convergence result that  $\beta \leq \limsup_{\alpha \searrow 0} \alpha V_\alpha(x)$  for all  $x \in \mathbb{R}^d$ .

We first show that  $\beta \leq \beta_*$ . For this, we employ the following assertion which is a special case of the Hardy–Littlewood theorem [39]: For any sequence  $\{a_n\}$  of non-negative real numbers, it holds that

$$\limsup_{\theta \nearrow 1} (1 - \theta) \sum_{n=1}^{\infty} \theta^n a_n \leq \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N a_n. \quad (\text{A.15})$$

Concerning this assertion, note that if the right hand side of the above display is finite then the set  $\{\frac{a_n}{n}\}$  is bounded. Therefore  $\sum_{n=1}^{\infty} \theta^n a_n$  is finite for every  $\theta < 1$ . Hence we can apply [39, Theorem 2.2] to obtain (A.15).

Fix  $x \in \mathbb{R}^d$ , and  $U \in \overline{\mathfrak{U}}_x$ . Define

$$a_n := \mathbb{E}_x^U \left[ \int_{n-1}^n \mathcal{R}(X_s, U_s) ds \right], \quad n \geq 1.$$

and let  $\theta = e^{-\alpha}$ . Applying (A.15), and with  $N$  running over the set of natural numbers, we obtain

$$\begin{aligned} \limsup_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_x^U \left[ \int_0^N \mathcal{R}(X_s, U_s) ds \right] &\geq \limsup_{\theta \nearrow 1} (1 - \theta) \sum_{n=1}^{\infty} \theta^n a_n \\ &\geq \limsup_{\alpha \searrow 0} (1 - e^{-\alpha}) \sum_{n=1}^{\infty} \mathbb{E}_x^U \left[ \int_{n-1}^n e^{-\alpha s} \mathcal{R}(X_s, U_s) ds \right] \\ &\geq \limsup_{\alpha \searrow 0} (1 - e^{-\alpha}) \mathbb{E}_x^U \left[ \int_0^\infty e^{-\alpha s} \mathcal{R}(X_s, U_s) ds \right] \\ &\geq \beta. \end{aligned} \quad (\text{A.16})$$

where we use the property that  $\limsup_{\alpha \searrow 0} \alpha V_\alpha(x) \geq \beta$ . Since  $U \in \overline{\mathfrak{U}}_x$  is arbitrary, (A.16) together with the definition of  $\beta_*$  imply that  $\beta \leq \beta_*$ . Note also that (A.16) implies that  $\mathfrak{U}_x^\alpha \subset \overline{\mathfrak{U}}_x$  for all  $\alpha \in (0, 1)$ .

Next, we prove the converse inequality. By (A.4) we have  $|\nabla V(x)| \leq c_0 (1 + |x|)$ . Therefore, since the Markov control  $v_* := -\nabla V(x)$  has at most linear growth, there exists a unique strong solution to (A.4) under the control  $v_*$ . Applying Itô's formula to (A.14), and using (3.1), we obtain

$$\mathbb{E}_x^{v_*} [V(X_{T \wedge \tau_R})] - V(x) + \mathbb{E}_x \left[ \int_0^{T \wedge \tau_R} \mathcal{R}[v_*](X_s) ds \right] = \beta \mathbb{E}_x [T \wedge \tau_R],$$

where  $\tau_R$  denotes the exit time from the ball of radius  $R > 0$  around 0. Since  $V$  is bounded from below and  $\tau_R \rightarrow \infty$  a.s., as  $R \rightarrow \infty$ , using Fatou's lemma for the integral on the left hand side, and then dividing by  $T$  and taking limits as  $T \rightarrow \infty$ , results in

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_x^{v_*} \left[ \int_0^T \mathcal{R}[v_*](X_s) ds \right] \leq \beta.$$

Thus  $\beta = \beta_*$ . Since  $\ell$  is inf-compact this also implies that the diffusion under the control  $v_*$  is positive recurrent, and by the ergodic theorem we obtain  $\beta_* = \int_{\mathbb{R}^d} \mathcal{R}[v_*](x) d\eta_*$ .  $\square$

*Proof of Theorem 1.4.* Part of the proof relies on results in [20]. These results require a Foster–Lyapunov type hypothesis which we verify next. Note that the operator  $F$  in [20] has a negative sign in the Laplacian so that  $\mathcal{A}[\varphi] = -F[\varphi]$ , where  $\mathcal{A}$  is the operator defined in (A.14). So, given that  $\ell$  is inf-compact,  $\varphi_0 = 0$  is an obvious choice to satisfy (A4) in [20]. Then of course  $-\mathcal{A}[\varphi_0] \rightarrow -\infty$  as  $|x| \rightarrow \infty$ . Note that Theorem 2.2 in [20] then asserts that  $V$  is bounded below in  $\mathbb{R}^d$ .

Next, consider  $\varphi_1 = -a_1\sqrt{\mathcal{V}}$  with  $a_1 := \inf_{\mathcal{K}^c} \frac{|\langle m, \nabla \mathcal{V} \rangle| \sqrt{\mathcal{V}}}{|\nabla \mathcal{V}|^2}$ , where  $\mathcal{K}$  is as in Hypothesis 1.1 (3) and  $\mathcal{V}$  is as in Lemma 2.3. Since  $\mathcal{V}$  agrees with  $\bar{\mathcal{V}}$  outside some compact set by Lemma 2.3, it follows by Hypothesis 1.1 (3) that  $a_1 > 0$ . Then we obtain

$$\begin{aligned} \frac{1}{2}\Delta\varphi_1 + \langle m, \nabla\varphi_1 \rangle - \frac{1}{2}|\nabla\varphi_1|^2 &= \frac{a_1}{4\sqrt{\mathcal{V}}} \Delta\mathcal{V} - \frac{a_1}{2\sqrt{\mathcal{V}}} \left( \langle m, \nabla\mathcal{V} \rangle + \frac{a_1\sqrt{\mathcal{V}}-1}{4\mathcal{V}} |\nabla\mathcal{V}|^2 \right) \\ &\geq \frac{a_1}{4\sqrt{\mathcal{V}}} (\Delta\mathcal{V} - \langle m, \nabla\mathcal{V} \rangle) \quad \text{on } \mathcal{K}^c. \end{aligned}$$

Thus, since  $\Delta\bar{\mathcal{V}}$  is bounded by Hypothesis 1.1 (3b), we obtain  $-\mathcal{A}[\varphi_1] \rightarrow -\infty$  as  $|x| \rightarrow \infty$ . It is also clear that  $\phi_0(x) - \phi_1(x) \rightarrow \infty$  as  $|x| \rightarrow \infty$ . Thus, Hypothesis (A.4)' in [20] is also satisfied. Therefore, as shown in [20, Theorem 2.1], there exists some critical value  $\lambda_*$  such that (1.12) has no solution for  $\beta > \lambda_*$ . Also by Theorem 2.2 and Corollary 2.3 in [20], the following hold: if  $V$  is a solution for  $\beta < \lambda_*$ , then under the control  $v = -\nabla V$ , the diffusion is transient. For  $\beta = \lambda_*$  there exists a unique solution  $V = V_*$  (up to an additive constant), and under the control  $v_* = -\nabla V_*$  the diffusion

$$X_t = X_0 + \int_0^t (m(X_s) + v(X_s)) \, ds + W_t, \quad t \geq 0,$$

is positive recurrent. It is clear then that Lemma A.3 implies that  $\lambda^* = \beta_*$ .

Part (a) of the theorem, was established in the proof of Lemma A.3; it also follows directly by [29, Lemma 5.1].

The uniqueness of the solution for  $\beta = \beta_*$  follows by the results in [23] discussed above, while the rest of the assertions in part (b) follow by Lemma A.3.

We now turn to part (c). Suppose  $\hat{v} \in \mathfrak{U}_{SSM}$  is an optimal stationary Markov control, and let  $\hat{\eta}$  denote the associated invariant probability measure. In particular,

$$\int_{\mathbb{R}^d} (\ell(x) + \frac{1}{2}|\hat{v}(x)|^2) \, \eta dx = \beta_*. \quad (\text{A.17})$$

It is straightforward to check from (A.3) that

$$\frac{1}{2}\Delta V_\alpha + \langle m + \hat{v}, \nabla V_\alpha \rangle + \frac{1}{2}|\hat{v}|^2 + \ell = \alpha V_\alpha + \frac{1}{2}|\nabla V_\alpha + \hat{v}|^2. \quad (\text{A.18})$$

By the very definition of  $\mathfrak{U}_{SSM}$ , the diffusion in (A.2) under the control  $\hat{v}$  is a strong Markov process. By applying Dynkin's formula to (A.18) we obtain

$$\begin{aligned} \mathbb{E}_x^{\hat{v}}[V_\alpha(X_{T \wedge \tau_R})] - V_\alpha(x) + \mathbb{E}_x^{\hat{v}} \left[ \int_0^{T \wedge \tau_R} (\ell + \frac{1}{2}|\hat{v}|^2)(X_s) \, ds \right] \\ = \mathbb{E}_x^{\hat{v}} \left[ \int_0^{T \wedge \tau_R} \left( \alpha V_\alpha + \frac{1}{2}|\nabla V_\alpha + \hat{v}|^2 \right) (X_s) \, ds \right]. \quad (\text{A.19}) \end{aligned}$$

The hypothesis that  $\hat{v}$  is optimal, and thus has finite control effort, implies by (A.9) that

$$\mathbb{E}_x^{\hat{v}} \left[ \sup_{0 \leq s \leq t} |X(s)| \right] < \infty \quad \forall x \in \mathbb{R}^d, \quad \forall t > 0. \quad (\text{A.20})$$

Therefore, since  $|V_\alpha(x)| \leq \tilde{c}_0(1 + |x|)$  for some constant  $\tilde{c}_0 = \tilde{c}_0(\alpha)$  by (A.4), then by dominated convergence implied by (A.20) we obtain  $\mathbb{E}_x^{\hat{v}}[V_\alpha(X_{T \wedge \tau_R})] \rightarrow \mathbb{E}_x^{\hat{v}}[V_\alpha(X_T)]$  as  $R \rightarrow \infty$ . Thus taking limits in (A.19) as  $R \rightarrow \infty$ , using monotone convergence for the integrals, it follows that (A.19) holds with  $T \wedge \tau_R$  replaced by  $T$ .

By (A.17) we obtain  $\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_x^{\hat{v}}[\ell(X_T)] = 0$ , which implies, in view of the estimate of  $V_\alpha$  in (A.4), that the same holds if we replace  $\ell$  by  $V_\alpha$ . Thus dividing (A.19) by  $T$  and taking limits as

$T \rightarrow \infty$ , using the property  $\frac{1}{T} \mathbb{E}_x^{\hat{v}}[V_\alpha(X_T)] \rightarrow 0$ , and also using the hypothesis that  $\hat{v}$  is optimal, we obtain

$$\beta_* \geq \int_{\mathbb{R}^d} \left( \alpha V_\alpha(x) + \frac{1}{2} |\nabla V_\alpha(x) + \hat{v}(x)|^2 \right) \hat{\eta}(dx). \quad (\text{A.21})$$

Since the subsequence  $\{\alpha_n\}$  over which (A.14) was obtained as a limit of (A.3) was arbitrary, and given the estimate of  $\nabla V_\alpha$  in (A.4), we have  $\lim_{\alpha \searrow 0} \alpha V_\alpha(x) = \beta_*$  for all  $x \in \mathbb{R}^d$ . Therefore, letting  $\alpha \searrow 0$  in (A.21) and applying again Fatou's lemma, we obtain

$$0 \geq \int_{\mathbb{R}^d} \frac{1}{2} |\nabla V + \hat{v}(x)|^2 \hat{\eta}(dx).$$

But since  $\hat{\eta}$  has a positive density, the equation above implies that  $\hat{v} = -\nabla V$  almost everywhere, thus completing the proof.  $\square$

## APPENDIX B. PROOFS OF THE RESULTS IN SECTION 1.4

We start with the proof of Lemma 1.16.

*Proof of Lemma 1.16.* Suppose that  $M$  has a number  $q$  of eigenvalues on the open right half complex plane. Using a similarity transformation we can transform  $M$  to a matrix of the form  $\text{diag}(M_1, -M_2)$  where  $M_1 \in \mathbb{R}^{(d-q) \times (d-q)}$  and  $M_2 \in \mathbb{R}^{q \times q}$  are Hurwitz matrices. So without loss of generality, we assume  $M$  has this form. Let  $S_1$  and  $S_2$  be the unique symmetric positive definite matrices solving the Lyapunov equations  $S_1 M_1 + M_1^\top S_1 = -I$  and  $S_2 M_2 + M_2^\top S_2 = -I$ , respectively. Extend these to symmetric matrices in  $\mathbb{R}^{d \times d}$  by defining  $\tilde{S}_1 = \text{diag}(S_1, 0)$  and  $\tilde{S}_2 = \text{diag}(0, S_2)$ , and also define, for  $\alpha > 0$ ,

$$\varphi_1(x) := e^{-\alpha \langle x, \tilde{S}_1 x \rangle}, \quad \varphi_2(x) := e^{-\alpha \langle x, \tilde{S}_2 x \rangle}, \quad \text{and } \varphi := 1 + \varphi_1 - \varphi_2.$$

Let  $T_1 = \text{diag}(I_{(d-q) \times (d-q)}, 0_{q \times q})$ , and similarly  $T_2 = \text{diag}(0_{(d-q) \times (d-q)}, I_{q \times q})$ . Then, with  $\mathcal{L}_v f(x) := \frac{1}{2} \Delta f(x) + \langle Mx + v(x), \nabla f(x) \rangle$ , we obtain

$$\begin{aligned} \mathcal{L}_v(1 - \varphi_2(x)) &= \alpha \varphi_2(x) \left( \text{trace}(\tilde{S}_2) - 2\alpha \langle x, \tilde{S}_2^2 x \rangle + |T_2 x|^2 + 2\langle v(x), \tilde{S}_2 x \rangle \right) \\ &\geq \alpha \varphi_2(x) \left( \text{trace}(\tilde{S}_2) + |T_2 x|^2 - \frac{1}{2} |T_2 x|^2 - 2\alpha \|\tilde{S}_2\|^2 |T_2 x|^2 - 2\|\tilde{S}_2\|^2 |v(x)|^2 \right) \\ &= \alpha \varphi_2(x) \left( \text{trace}(S_2) + \left( \frac{1}{2} - 2\alpha \|\tilde{S}_2\|^2 \right) |T_2 x|^2 - 2\|\tilde{S}_2\|^2 |v(x)|^2 \right). \end{aligned} \quad (\text{B.1})$$

For the inequality in (B.1) we use

$$\begin{aligned} 2\langle v(x), \tilde{S}_2 x \rangle &= 2\langle \tilde{S}_2 v(x), T_2 x \rangle \geq -\left| \frac{T_2 x}{\sqrt{2}} \right|^2 - |\sqrt{2}\tilde{S}_2 v(x)|^2 \\ &\geq -\frac{1}{2} |T_2 x|^2 - 2\|\tilde{S}_2\|^2 |v(x)|^2. \end{aligned}$$

Using the analogous inequality for  $\mathcal{L}_v \varphi_2(x)$  and combining the equations we obtain

$$\begin{aligned} \mathcal{L}_v \varphi(x) &\geq \alpha e^{-\alpha \langle x, \tilde{S}_1 x \rangle} \left( -\text{trace}(S_1) + \left( \frac{1}{2} + 2\alpha \|\tilde{S}_1\|^2 \right) |T_1 x|^2 - 2\|\tilde{S}_1\|^2 |T_1 v(x)|^2 \right) \\ &\quad + \alpha e^{-\alpha \langle x, \tilde{S}_2 x \rangle} \left( \text{trace}(S_2) + \left( \frac{1}{2} - 2\alpha \|\tilde{S}_2\|^2 \right) |T_2 x|^2 - 2\|\tilde{S}_2\|^2 |T_2 v(x)|^2 \right) \\ &\geq \alpha \left( -\text{trace}(S_1) + e^{-\alpha \langle x, S x \rangle} \left( \frac{1}{2} - 2\alpha \|S\|^2 \right) |x|^2 - 2\|S\|^2 |v(x)|^2 \right), \end{aligned} \quad (\text{B.2})$$

with  $S := \text{diag}(S_1, S_2)$ .

Using Itô's formula on (B.2), dividing by  $\alpha$ , and also using the fact that  $\varphi \geq 0$  and  $\|\varphi\|_\infty = 2$ , we obtain

$$\mathbb{E}_x \left[ \int_0^T \left( -\text{trace}(S_1) + e^{-\alpha \langle X_t, S X_t \rangle} \left( \frac{1}{2} - 2\alpha \|S\|^2 \right) |X_t|^2 - 2\|S\|^2 |v(X_t)|^2 \right) dt \right] \leq \frac{2}{\alpha}.$$

Dividing by  $T$ , letting  $T \nearrow \infty$  and rearranging terms, we conclude that  $e^{-\alpha \langle x, Sx \rangle} |x|^2$  is integrable with respect to invariant probability measure  $\mu_v$  under the control  $v$  for any  $\alpha < \frac{1}{4\|S\|^2}$ , and the following bound holds

$$\int_{\mathbb{R}^d} e^{-\alpha \langle x, Sx \rangle} |x|^2 \mu_v(dx) \leq \frac{\text{trace}(S_1)}{\left(\frac{1}{2} - 2\alpha \|S\|^2\right)} + \frac{2\|S\|^2}{\left(\frac{1}{2} - 2\alpha \|S\|^2\right)} \int_{\mathbb{R}^d} |v(x)|^2 \mu_v(dx).$$

Taking limits as  $\alpha \searrow 0$ , using monotone convergence, we obtain

$$\int_{\mathbb{R}^d} |x|^2 \mu_v(dx) \leq \text{trace}(S_1) + 4\|S\|^2 \int_{\mathbb{R}^d} |v(x)|^2 \mu_v(dx).$$

The proof is complete.  $\square$

*Proof Theorem 1.18.* It is well known that there exists at most one symmetric matrix  $Q$  satisfying (1.26)-(1.27) [12, Theorem 3, p. 150]. For  $\kappa > 0$ , consider the ergodic control problem of minimizing

$$J_\kappa(v) := \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \int_0^T \left( \kappa |X_s|^2 + \frac{1}{2} |v(X_s)|^2 \right) ds \right], \quad (\text{B.3})$$

over  $v \in \bar{\mathfrak{U}}_{\text{SSM}}$ , subject to the linear controlled diffusion

$$X_t = X_0 + \int_0^t (MX_s + v(X_s)) ds + W_t, \quad t \geq 0. \quad (\text{B.4})$$

As is also well known, an optimal stationary Markov control for this problem takes the form  $v(x) = -Q_\kappa x$ , where  $Q_\kappa$  is the unique positive definite symmetric solution to the matrix Riccati equation

$$Q_\kappa^2 - M^\top Q_\kappa - Q_\kappa M = 2\kappa I. \quad (\text{B.5})$$

Moreover,  $Q_\kappa$  has the following property: Consider a deterministic linear control system  $\dot{x}(t) = Mx(t) + u(t)$ , with  $x, u \in \mathbb{R}^d$ , and initial condition  $x(0) = x_0$ . Let  $\mathcal{U}$  denote the space of controls  $u$  satisfying  $\int_0^T |u(t)|^2 dt < \infty$  for all  $T > 0$ , and  $\phi_t^u(x_0)$  denote the solution of the differential equation under a control  $u \in \mathcal{U}$ . Then

$$\langle x_0, Q_\kappa x_0 \rangle = \min_{u \in \mathcal{U}} \int_0^\infty \left( |u(t)|^2 + 2\kappa |\phi_t^u(x_0)|^2 \right) dt. \quad (\text{B.6})$$

For these assertions, see [12, Theorem 1, p. 147].

On the other hand,  $\Psi_\kappa(x) = \frac{1}{2} \langle x, Q_\kappa x \rangle$  is a solution of the associated HJB equation

$$\frac{1}{2} \Delta \Psi_\kappa(x) + \min_{u \in \mathbb{R}^d} \left[ \langle Mx + u, \nabla \Psi_\kappa(x) \rangle + \frac{1}{2} |u|^2 \right] + \kappa |x|^2 = \frac{1}{2} \text{trace}(Q_\kappa). \quad (\text{B.7})$$

The HJB equation (B.7) characterizes the optimal cost, i.e.,

$$\inf_{v \in \bar{\mathfrak{U}}_{\text{SSM}}} J_\kappa(v) = \frac{1}{2} \text{trace}(Q_\kappa).$$

Recall Definition 1.17. Since the stationary probability distribution of (B.4) under the control  $v(x) = -Q_\kappa x$  is Gaussian, it follows by (B.3) that  $G = Q_\kappa$  minimizes

$$\tilde{\mathcal{J}}_{G,\kappa}(M) := \kappa \text{trace}(\Sigma_G) + \frac{1}{2} \text{trace}(G \Sigma_G G^\top)$$

over all matrices  $G \in \mathcal{G}(M)$ , where  $\Sigma_G$  is as in (1.24) (note that  $\tilde{\mathcal{J}}_{G;0}(M) = \mathcal{J}_G(M)$  which is the right hand side of (1.25)). Combining this with (B.7) we have

$$\inf_{G \in \mathcal{G}(M)} \tilde{\mathcal{J}}_{G;\kappa}(M) = \tilde{\mathcal{J}}_{Q_\kappa;\kappa}(M) = \frac{1}{2} \text{trace}(Q_\kappa). \quad (\text{B.8})$$

By Lemma 1.16 we have

$$\begin{aligned} \text{trace}(\Sigma_{Q_\kappa}) &\leq \tilde{C}_0(1 + \mathcal{J}_{Q_\kappa;\kappa}(M)) \\ &= \tilde{C}_0\left(1 + \frac{1}{2} \text{trace}(Q_\kappa)\right). \end{aligned} \quad (\text{B.9})$$

It also follows by (B.6) that  $Q_{\kappa'} - Q_\kappa$  is nonnegative definite if  $\kappa' \geq \kappa$ . Therefore  $Q_\kappa$  has a unique limit  $Q$  as  $\kappa \searrow 0$ . It is evident that  $Q$  is nonnegative semidefinite, and (B.5) shows that it satisfies (1.26). Since  $\text{trace}(\Sigma_{Q_\kappa})$  is bounded by (B.9), it follows that  $\Sigma_{Q_\kappa}$  converges along some subsequence  $\kappa_n \searrow 0$  to a symmetric positive semidefinite matrix  $\Sigma$ . Thus (1.27) holds. However, (1.27) implies that  $\Sigma$  is invertible, and therefore, it is positive definite. In turn, (1.27) implies that  $M - Q$  is Hurwitz.

Since  $v_G(x) = -Gx$ ,  $G \in \mathcal{G}(M)$ , is in general suboptimal for the criterion  $J_\kappa(v)$ , applying Lemma 1.16 once more, we obtain

$$\mathcal{J}_{Q_\kappa}(M) \leq \tilde{\mathcal{J}}_{Q_\kappa;\kappa}(M) \leq \kappa \tilde{C}_0(1 + \mathcal{J}_G(M)) + \mathcal{J}_G(M) \quad \forall G \in \mathcal{G}(M).$$

Therefore, we have

$$\mathcal{J}_*(M) \leq \tilde{\mathcal{J}}_{Q_\kappa;\kappa}(M) \leq \kappa \tilde{C}_0(1 + \mathcal{J}_*(M)) + \mathcal{J}_*(M),$$

and taking limits as  $\kappa \searrow 0$ , this implies by (B.8) that  $\mathcal{J}_*(M) = \frac{1}{2} \text{trace}(Q)$ .

It remains to show that  $\Lambda^+(M) = \frac{1}{2} \text{trace}(Q)$ . Let  $T$  be a unitary matrix such that  $\tilde{Q} := TQT^\top$  takes the form  $\tilde{Q} = \text{diag}(0, \tilde{Q}_2)$ , with  $\tilde{Q} \in \mathbb{R}^{q \times q}$  a positive definite matrix. Write the corresponding block structure of  $TMT^\top$  as

$$\tilde{M} := TMT^\top = \begin{pmatrix} \tilde{M}_{11} & \tilde{M}_{12} \\ \tilde{M}_{21} & \tilde{M}_{22} \end{pmatrix},$$

with  $\tilde{M}_{22} \in \mathbb{R}^{q \times q}$ . Since  $M^\top Q + QM = Q^2$ , we obtain  $\tilde{M}^\top \tilde{Q} + \tilde{Q} \tilde{M} = \tilde{Q}^2$ , and block multiplication shows that  $\tilde{Q}_2 \tilde{M}_{21} = 0$ , which implies that  $\tilde{M}_{21} = 0$ . Since  $M - Q$  is similar to  $\tilde{M} - \tilde{Q}$  the latter must be Hurwitz, which implies that  $\tilde{M}_{11}$  is Hurwitz. By block multiplication we have

$$\tilde{M}_{22}^\top \tilde{Q}_2 + \tilde{Q}_2 \tilde{M}_{22} = \tilde{Q}_2^2. \quad (\text{B.10})$$

Since  $\tilde{Q}_2$  is positive definite, the matrix  $-\tilde{M}_{22}$  is Hurwitz by the Lyapunov theorem. Thus  $\Lambda^+(M) = \text{trace}(\tilde{M}_{22})$ . Therefore, since  $\tilde{Q}_2$  is invertible, we obtain by (B.10) that

$$\begin{aligned} \text{trace}(Q) &= \text{trace}(\tilde{Q}_2) = \text{trace}(\tilde{M}_{22}^\top + \tilde{M}_{22}) \\ &= 2 \text{trace}(\tilde{M}_{22}) = 2\Lambda^+(M). \end{aligned}$$

This proves part (a).

Now let  $\hat{v} \in \bar{\mathfrak{U}}_{\text{SSM}}$  be any control. Let  $\bar{V}(x) = \frac{1}{2}\langle x, Qx \rangle$ . Then  $\bar{V}$  satisfies (B.7) with  $\kappa = 0$ , and we have

$$\frac{1}{2}\Delta\bar{V}(x) + \langle Mx + \hat{v}(x), \nabla\bar{V}(x) \rangle + \frac{1}{2}|\hat{v}(x)|^2 = \frac{1}{2}\text{trace}(Q) + \frac{1}{2}|Qx + \hat{v}(x)|^2. \quad (\text{B.11})$$

Applying Itô's formula to (B.11), and using the fact that  $\mu_{\hat{v}}$  has finite second moments as shown in Lemma 1.16, and  $\bar{V}$  is quadratic, a standard argument gives

$$\int_{\mathbb{R}^d} \left( \frac{1}{2}|\hat{v}(x)|^2 - \frac{1}{2}|Qx + \hat{v}(x)|^2 \right) \mu_{\hat{v}}(dx) = \frac{1}{2} \text{trace}(Q). \quad (\text{B.12})$$

Thus  $\int_{\mathbb{R}^d} \frac{1}{2}|\hat{v}(x)|^2 \mu_{\hat{v}}(dx) \geq \frac{1}{2} \text{trace}(Q) = \mathcal{J}_*(M)$ . Hence (1.28) holds.

Suppose  $\hat{v}$  is optimal, i.e., attains the infimum in (1.28). By (B.12), we obtain

$$\lim_{\kappa \searrow 0} \int_{\mathbb{R}^d} |Qx + \hat{v}(x)|^2 \mu_{\hat{v}}(dx) = 0.$$

Therefore, since  $\mu_{\hat{v}}$  has a positive density, it holds that  $\hat{v}(x) = -Qx$  a.e. in  $\mathbb{R}^d$ . This completes the proof of part (b).

We have shown that  $\bar{V}(x) = \frac{1}{2}\langle x, Qx \rangle$  satisfies (1.29) with  $\bar{\beta} = \Lambda^+(M)$  and the associated process is positive recurrent. Therefore, as in the proof of Theorem 1.4 for a bounded  $m$ , part (c) follows by Theorems 2.1–2.2 and Corollary 2.3 in [20]. Note that Hypothesis (A4) in [20] is easily satisfied for the linear problem. Since  $M$  is exponentially dichotomous, then as seen in the proof of Theorem 2.2, there exists symmetric matrices  $S$  and  $\hat{S}$ , with  $\hat{S}$  positive definite such that  $M^T S + S M = \hat{S}$ . Consider the function  $\varphi_0(x) := a \langle x, Sx \rangle$ , with  $a := \frac{1}{4}(\|\hat{S}^{-1}\| \|S\|)^{-1}$ . Then

$$\begin{aligned} \bar{\mathcal{A}}[\varphi_0](x) &:= \frac{1}{2} \Delta \varphi_0(x) + \langle Mx, \nabla \varphi_0(x) \rangle - \frac{1}{2} |\nabla \varphi_0(x)|^2 \\ &= a \operatorname{trace} S + a \langle x, \hat{S}x \rangle - 2a^2 |Sx|^2 \\ &> \frac{a}{2} (2 \operatorname{trace} S - \langle x, \hat{S}x \rangle). \end{aligned}$$

Thus  $\bar{\mathcal{A}}[\varphi_a](x) \rightarrow \infty$  as  $|x| \rightarrow \infty$ . This completes the proof.  $\square$

#### ACKNOWLEDGMENT

The authors are indebted to the anonymous referees for their constructive comments and suggestions. This work was initiated during Vivek Borkar's visit to the Department of Electrical Engineering, Technion, supported by Technion. Thanks are due to Prof. Rami Atar for suggesting the problem as well as for valuable discussions. The work of Ari Arapostathis was supported in part by the Office of Naval Research through grant N00014-14-1-0196. The work of Anup Biswas was supported in part by an award from the Simons Foundation (# 197982) to The University of Texas at Austin and in part by the Office of Naval Research through the Electric Ship Research and Development Consortium. This research of Anup Biswas was also supported in part by an INSPIRE faculty fellowship. The work of Vivek Borkar was supported in part by a J. C. Bose Fellowship from the Department of Science and Technology, Government of India.

#### REFERENCES

- [1] A. Arapostathis, V. S. Borkar, and M. K. Ghosh. *Ergodic control of diffusion processes*, volume 143 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2011.
- [2] A. Bensoussan and J. Frehse. On Bellman equations of ergodic control in  $\mathbf{R}^n$ . *J. Reine Angew. Math.*, 429:125–160, 1992.
- [3] A. Bensoussan and J. Frehse. *Regularity results for nonlinear elliptic systems and applications*, volume 151 of *Applied Mathematical Sciences*. Springer-Verlag, Berlin, 2002.
- [4] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani. A theory of stochastic resonance in climatic change. *SIAM J. Appl. Math.*, 43(3):565–478, 1983.
- [5] N. Berglund and B. Gentz. Metastability in simple climate models: pathwise analysis of slowly driven Langevin equations. *Stoch. Dyn.*, 2(3):327–356, 2002.
- [6] N. Berglund and B. Gentz. *Noise-induced phenomena in slow-fast dynamical systems. A sample-paths approach*. Probability and its Applications (New York). Springer-Verlag London, Ltd., London, 2006.
- [7] A. G. Bhatt and V. S. Borkar. Occupation measures for controlled Markov processes: characterization and optimality. *Ann. Probab.*, 24(3):1531–1562, 1996.
- [8] A. Biswas and V. S. Borkar. Small noise asymptotics for invariant densities for a class of diffusions: a control theoretic view. *J. Math. Anal. Appl.*, 360(2):476–484, 2009. Erratum at arXiv:1107.2277.
- [9] B. Z. Bobrovsky, M. M. Zakai, and O. Zeitouni. Error bounds for the nonlinear filtering of signals with small diffusion coefficients. *IEEE Trans. Inform. Theory*, 34(4):710–721, 1988.
- [10] V. I. Bogachev, M. Röckner, and S. V. Shaposhnikov. On positive and probability solutions of the stationary Fokker-Planck-Kolmogorov equation. *Dokl. Akad. Nauk*, 444(3):245–249, 2012.

- [11] V. S. Borkar. *Optimal control of diffusion processes*, volume 203 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1989.
- [12] R. W. Brockett. *Finite dimensional linear systems*. John Wiley & Sons, 1970.
- [13] S. Cerrai and M. Röckner. Large deviations for invariant measures of stochastic reaction-diffusion systems with multiplicative noise and non-Lipschitz reaction term. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(1):69–105, 2005.
- [14] M. V. Day. Recent progress on the small parameter exit problem. *Stochastics*, 20(2):121–150, 1987.
- [15] E. Fedrizzi and F. Flandoli. Hölder flow and differentiability for SDEs with nonregular drift. *Stoch. Anal. Appl.*, 31(4):708–736, 2013.
- [16] J. Feng, M. Forde, and J.-P. Fouque. Short-maturity asymptotics for a fast mean-reverting Heston stochastic volatility model. *SIAM J. Financial Math.*, 1(1):126–141, 2010.
- [17] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, New York, second edition, 1998.
- [18] I. Gyöngy and N. Krylov. Existence of strong solutions for Itô’s stochastic equations via approximations. *Probab. Theory Related Fields*, 105(2):143–158, 1996.
- [19] S. Herrmann, P. Imkeller, I. Pavlyukevich, and D. Peithmann. *Stochastic resonance*, volume 194 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2014.
- [20] N. Ichihara. Recurrence and transience of optimal feedback processes associated with Bellman equations of ergodic type. *SIAM J. Control Optim.*, 49(5):1938–1960, 2011.
- [21] N. Ichihara. Large time asymptotic problems for optimal stochastic control with superlinear cost. *Stochastic Process. Appl.*, 122(4):1248–1275, 2012.
- [22] N. Ichihara. The generalized principal eigenvalue for Hamilton-Jacobi-Bellman equations of ergodic type. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 32(3):623–650, 2015.
- [23] N. Ichihara and S.-J. Sheu. Large time behavior of solutions of Hamilton-Jacobi-Bellman equations with quadratic nonlinearity in gradients. *SIAM J. Math. Anal.*, 45(1):279–306, 2013.
- [24] N. V. Krylov and M. Röckner. Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Related Fields*, 131(2):154–196, 2005.
- [25] V. Kučera. A contribution to matrix quadratic equations. *IEEE Trans. Automatic Control*, AC-17(3):344–347, 1972.
- [26] T. G. Kurtz and R. H. Stockbridge. Existence of Markov controls and characterization of optimal Markov controls. *SIAM J. Control Optim.*, 36(2):609–653, 1998.
- [27] John F. Lindner, Matthew Bennett, and Kurt Wiesenfeld. Potential energy landscape and finite-state models of array-enhanced stochastic resonance. *Phys. Rev. E*, 73:031107, Mar 2006.
- [28] K. Mårtensson. On the matrix Riccati equation. *Information Sci.*, 3:17–49, 1971.
- [29] G. Metafune, D. Pallara, and A. Rhandi. Global properties of invariant measures. *J. Funct. Anal.*, 223(2):396–424, 2005.
- [30] K. R. Meyer. Energy functions for Morse Smale systems. *Amer. J. Math.*, 90:1031–1040, 1968.
- [31] F. Moss. Stochastic resonance: From the ice ages to the monkey’s ear. In George H. Weiss, editor, *Contemporary Problems in Statistical Physics*, chapter 5, pages 205–253. SIAM, Philadelphia, 1994.
- [32] E. Olivieri and M. E. Vares. *Large deviations and metastability*, volume 100 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005.
- [33] N. I. Portenko. *Generalized diffusion processes*, volume 83 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1990. Translated from the Russian by H. H. McFaden.
- [34] D. W. Repperger and K. A. Farris. Stochastic resonance—a nonlinear control theory interpretation. *Internat. J. Systems Sci.*, 41(7):897–907, 2010.
- [35] D. F. Russell, Wilkens L. A., and Moss F. Use of behavioural stochastic resonance by paddle fish for feeding. *Nature*, 402:291–294, 1999.
- [36] Z. Schuss. *Theory and applications of stochastic differential equations*. John Wiley & Sons, Inc., New York, 1980.
- [37] S. J. Sheu. Asymptotic behavior of the invariant density of a diffusion Markov process with small diffusion. *SIAM J. Math. Anal.*, 17(2):451–460, 1986.
- [38] S. Smale. On gradient dynamical systems. *Ann. of Math. (2)*, 74:199–206, 1961.
- [39] R. Sznajder and J. A. Filar. Some comments on a theorem of Hardy and Littlewood. *J. Optim. Theory Appl.*, 75(1):201–208, 1992.
- [40] J. C. Willems. Least squares stationary optimal control and the algebraic Riccati equation. *IEEE Trans. Automatic Control*, AC-16:621–634, 1971.
- [41] X. Wu, Z.-P. Jiang, D. W. Repperger, and Y. Guo. Enhancement of stochastic resonance using optimization theory. *Commun. Inf. Syst.*, 6(1):1–18, 2006.
- [42] Y. Yang, Z.-P. Jiang, B. Xu, and D. W. Repperger. An investigation of two-dimensional parameter-induced stochastic resonance and applications in nonlinear image processing. *J. Phys. A*, 42(14):145207, 9, 2009.

- [43] O. Zeitouni and M. Zakai. On the optimal tracking problem. *SIAM J. Control Optim.*, 30(2):426–439, 1992.
- [44] X. Zhang. Strong solutions of SDES with singular drift and Sobolev diffusion coefficients. *Stochastic Process. Appl.*, 115(11):1805–1818, 2005.